1-1-2004

Phylogenetic study of glycogen synthesizing enzymes in Synechocystis

Wei He
Iowa State University

Phylogenetic study of glycogen synthesizing enzymes in *Synechocystis*

by

Wei He

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Plant Physiology

Program of Study Committee:
Martin Spalding, Major Professor
Jonathan Wendel
Paul Scott

Iowa State University

Ames, Iowa

2004

Graduate College
Iowa State University

This is to certify that the master's thesis of

Wei He

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

INTRODUCTION

Glycogen is a major reserve carbohydrate in both prokaryotes and eukaryotes, and starch is a major reserve carbohydrate only in photosynthetic eukaryotes. Synthesis of both glycogen and starch involves at least glycogen (starch) synthase and branching enzyme, as well as perhaps debranching enzyme. Extensive biochemical and genetic research has been carried out on glycogen and starch biosynthesis in bacteria and plants. However, it is not yet understood what controls the structure of these two glucans to result in such different forms or how glycogen synthesis (certainly more ancient) gave rise to starch synthesis. This phylogenetic study on the evolutionary relationships of bacterial glycogen synthases with plant starch synthases and bacterial glycogen debranching enzymes with plant isoamylases, combined with the published studies of the origin of plastids in higher plants from an endosymbiotic event, reveal a close evolutionary relationship between glycogen and starch biosynthesis and suggest a potential origin for starch biosynthesis.

LITERATURE REVIEW

Starch is the storage form of glucose in eukaryotic photosynthetic organisms. Glycogen is its equivalent in bacteria, cyanobacteria, yeast, animals and humans. Both starch and glycogen are glucose polymers in which the monomers are connected by α-1,4 glycosidic linkages and contain α-1,6 branch points. Even though starch and glycogen are chemically very similar, there are many differences between the two, especially with regard to structure (Ball et al., 2003). Starch has a clustered branching pattern, whereas glycogen has a more random or homogeneous branching pattern. Starch generally is a much larger molecule (amylopectin: 10^7-10^9 Da, amylose: 10^5-10^7 Da), has higher degree of polymerization (DP; number of linked glucose units in each branch), and has branches spaced farther apart (10-12 glucose units for starch vs 8-10 glucose units in glycogen).

Starch is synthesized exclusively inside the plastid compartment of eukaryotic photosynthetic cells as an integral part of metabolism. It forms large, semi-crystalline, insoluble granules composed of 10-20% amylose and 80-90% amylopectin. Amylopectin is a branched glucan composed of α-1, 4-linked D-glucopyranose connected by α-1,6-linked branch points in a heterogeneous pattern, in which highly branched regions alternate with relatively unbranched regions. This clustered branching pattern allows neighboring chains to interact to form parallel arrays of double helices in a complex ordered array (Jenkins et al., 1993). Amylose is a linear α-1, 4-linked D-glucopyranose, arranged inside the starch granule and usually exists in a random coil form if not complexed by lipids or fatty acids (Smith, 2001). However, amylose also can interact with fatty acids or lysophospholipids and form single helices where the interior is hydrophobic, because of the hydrogen atoms (Morrison et al., 1984).

The enzymes necessary for starch synthesis and structure determination are starch synthase (SS), starch branching enzyme (BE) and starch debranching enzyme (DBE) (James, et al., 2003). Using ADP-glucose (ADPG) as substrate, SS catalyzes formation of α-1, 4 linkages. There are four or five types of soluble starch synthases, SSI – SSV and one or two types of granule bound starch synthases, gbSSI and gbSSII, and these types are defined based on sequence homology, molecular mass and antigenic properties (Marshall, et al., 1996; Martin, et al., 1995). It is not clear how all these starch synthases cooperate to form a starch granule, but some characteristics are known for the different synthases. Maize SSI has been shown to

be catalytically active only with glucan chains shorter than dp<10 and preferentially extends shorter chained (dp<5) malto-oligosaccharides (Commuri et al., 2001). SSII from pea and *Chlamydomonas* has the unique function of elongating very short chains to generate chains that is the basis of clusters of amylopectin (Smith et al. 1997). In monocots SSII can be broken into two subclasses, SSIIa and SSIIb. Monocot SSIIa and dicot SSII gene play similar roles in starch biosynthesis, but without identified mutants, the function of SSIIb in endosperm starch biosynthesis is not known (Ball et al., 2003). SSII and SSIII have different but not independent effects on the synthesis of amylopectin, in that the product of one may be the substrate of the other (Smith 2001). SSIII mutants from *Chlamydomonas* (sta3) (Fontaine T et al., 1993) and maize (du1) (Gao et al., 1996) show reduced starch synthesis and different chain-length distribution from wildtype. Little is known about the role of SSIV. SSV is considered by some to be the same as SSIV, but no protein product or genetic evidence has been identified about this (Ball et al., 2003). The function of gbSSI is known to include elongating amylose, because in mutants defective in gbSSI, such as the waxy mutants of cereals or other mutations in pea, potato, etc., there is no amylose formed, so only amylopectin is found in the starch granules (Smith 2001). In wheat GBSSII expresses in pericarp, leaf, and culm tissues and GBSSI expresses only in endosperm. This tissue-specificity affects the amylose content of starch granules (Vrinten et al., 2000). In a phylogenetic analysis conducted on SSI, SSII, SSIII, SSIV, gbSSI and gbSSIb sequences from Maize, Wheat, Rice, Potato, Cowpea, *Arabidopsis* and pea (Ball et al. 2003), SSIII and SSIV formed one monophyletic group and SSI, SSII, gbSSI and gbSSIb formed a separate monophyletic group. Within one of the main clades, SSI, SSII and gbSS formed clearly defined subclades, with monocot and dicot forms of each being further distinguished. Within the other main clades, SSIII and SSIV also formed distinct subclades. The separation between the two main clades suggests the possibility of at least slightly different evolutionary origins for the two main groups of starch synthases.

BE hydrolyzes α-1,4-bonds and creates α-1, 6-bonds by transferring reducing ends to the C6 of a glucose unit within the glucan chain. There are two types of BE, BEI and BEII, and two isoforms of BEII, BEIIa and BEIIb. Those isoforms vary in the temporal and spatial patterns of expression (James et al., 2003). In maize, BEI is distinguishable from BEII in substrate

specificity and transferability of α-1, 4-linked linear chains. Maize BEI has a greater activity in branching amylose, whereas maize BEII has greater activity in branching amylopectin (Guan, et al., 1993). Maize BEI and BEII also have differing preferences in the length of the chains transferred, with BEI preferentially transferring longer (~12 glucose units) and BEII transferring shorter (~ 6 glucose units) chains, respectively. In pea, SBEI and SBEII are expressed at different times, with SBEI expressed at the early stages of embryo development and SBEII expressed in the older embryos (Mizuno et al., 2001).

DBE hydrolyzes α-1, 6 bonds in α-1, 4/ α-1, 6-liked glucans. Although DBEs clearly play a role in starch hydrolysis and catabolism, the importance of DBE for normal starch accumulation in plants was discovered when it was observed that mutants with decreased or no DBE activity, such as the Sugary1 (su1) mutant of maize, accumulate soluble, homogenously branched phytoglycogen (Back et al., 1966). The su1 gene encodes a DBE (James et al., 1995). The sugary mutant in rice, the notch2 mutant in barley (Fujita et al., 1999; Kubo et al., 1999; Burton et al., 2002), the Sta7 mutant in *Chlamydomonas reinhardtii* (Mouille et al., 1996) and a DBE deletion in *Arabidopsis* (Zeeman et al., 1998) all show phytoglycogen accumulation and reduced starch synthesis. There are two types of DBEs, isoamylase (IA) type and pullulanase type, differing in their substrate preference. Although having related primary amino acid sequences, those two DBE types differ in their distinct motifs (James et al., 1995; Beatty et al., 1999). The IA-type DBEs are the type apparently required for starch synthesis (Hussain et al., 2003), although there is evidence that pullulanase types can also contribute to starch synthesis, at least in mutants already deficient in isoamylase (Dinges et al., 2003). Three isoforms of IA, stisa1, stisa2 and stisa3, have been identified in potato (Hussain et al., 2003), *Arabidopsis* and rice (Arabidopsis Genome Initiative, 2000; Goff et al., 2002; Yu et al., 2002). The three have different catalytic specificities. Stisa1 and stisa2 are associated as a multimeric enzyme that functions to debranch soluble glucan, serving a central role in starch synthesis (James et al., 1995; Nakamura et al., 1996; Zeeman et al., 1998; Kubo et al., 1999; Burton et al., 2002). Stisa3 is similar to the STA7 gene in Chlamydomonas and can also interact with stisa2, forming a separate complex active in debranching substrates with different specificity from the stisa1/stisa2 complex (Hussain et al., 2003). Phylogenetic studies of IA isoform types from

Chlamydomonas, Arabidopsis, potato, barley, wheat and maize show that the three isoforms appear to be present in both monocots and dicots and fall into structurally different isoform classes (Hussain et al., 2003), which suggest that they may represent evolutionarily distinct isoforms.

Glycogen is an α-D-(1-4)-glucan with 10% α-(1-6)-linked branches. Major enzymes that function in glycogen synthesis are glycogen synthase (GS), glycogen branching enzyme (GBE), and perhaps glycogen debranching enzyme (GDBE). Using UDP-glucose (UDPG) in mammalian and fungal cells or ADPG in bacterial cells, GS catalyzes a reaction very similar to starch synthase, forming linear glucan chains with α-1, 4 linkages. Although most bacteria and cyanobacteria only have a single isoform of GS, *Synechocystis* sp. PCC6803 has been shown to have two isoforms of GS, based on genomic DNA sequences and zymogram analysis. These two isoforms are both involved in glycogen biosynthesis (Yoo et al., in preparation).

GBEs also carry out reactions very similar to those of starch BE, i.e., they hydrolyze α-1, 4-bonds and create α-1, 6-bonds by transferring reducing ends to a C6 of one of the glucose units in the glucan chain. Thus GBEs are essential to the formation of structurally normal glycogen, so GBE deletion mutants synthesize essentially linear glucans rather than the normal branched glucans (Yoo et al., 2002). Unlike the case of multiple starch BE isoforms in photosynthetic eukaryotes, there appears to be only a single form of GBE in bacteria and cyanobacteria.

Bacteria and cyanobacteria also have both pullulanase-like and isoamylase-like isoforms of GDBEs. However, little is known about the functions of these isoforms in prokaryotes. There is no evidence to suggest that either form plays an essential function in glycogen biosynthesis.

It is well established that plastids, the light-harvesting and starch synthesizing organelles of photosynthetic eukaryotes, arose from an ancient symbiosis between a eukaryotic cell and a cyanobacterium (Archibald, 2002). A single endosymbiotic event with a cyanobacterium apparently gave rise to plant plastids, based on the gene sequence of mitochondria (Burger, 1999), plastids (Douglas, 1999; Turmel, 1999), and nuclei (Baldauf, 2000; Moreira, 2000) of plants. After comparing 24,990 proteins encoded in the Arabidopsis genome to proteins from

three cyanobacteria genomes, 16 other prokaryotic reference genomes and yeast, the cyanobacterial ancestor of the plastids was concluded to be more closely related to *Nostoc punctiforme* than to *Prochlorococcus marinus* or *Synechocystis* (Martin, 2002). Subsequent to the initial endosymbiotic event, genes in pre-plastids apparently were retained, lost, transferred to the nucleus or duplicated, with the nuclear form or copy of the gene producing a functional product either in the cytosol or targeted to other compartments, including the plastids (Raven, 2003). This phylogenetic comparison of SS genes with GS genes and SDBE genes with GDBE genes suggests a possible relationship between the glycogen synthesis genes of the original endosymbiont and the origin of starch synthesis.

MATERIALS AND METHODS

Protein Sequences Dataset and Alignment

Protein sequences were extracted from NCBI and the complete genomes of *Synechocystis* sp.PCC6903, *Anabaena* sp. PC7120, *Thermosynechococcus elongates* BP-1, *Gloeobacter violaceus* PCC7421, *Prochlorococcus marinus* SS 120, *Prochloroccus marinus* MED4, *Prochlorococcus marinus* MIT9313, *Synechococcus* sp. WH8102, *Chlorobium tepidum* TLS, *Rhodopseudomonas palustris* CGA009 from the Genome Database for Cyanobacteria (Kaneko et al., 1997) and *Crocosphaera watsonii* WH8501, *Trichodesmium erythraeum*, *Chlamydomonas reinhardtii* (Shrager et al. 2003) from from JGI((Doe Joint Genome Institute, US Department of Energy).

Protein sequences were obtained by performing pBlastp (Altschul et al., 1990) to recruit sequences using *Synechocystis* GSI and type I IA-type GDBE protein sequences as the query sequence to obtain the best match protein sequences in NCBI database. All the protein sequences with bit score higher than 1e-20 were recruited and compiled as a .txt file. Then redundant sequences were eliminated and representative sequences selected for each phylum according to the NCBI taxonomy database. It was sometimes necessary to correct some of the annotation inconsistency according to the sources of the submission. This process resulted in development of a crude protein sequence dataset with which to work.

ClustalX, an updated version of ClustalW, was used as the alignment software (Thompson et al. 1994). It recognizes .txt, .aln and .fas file and produce .fas and .nxs output file. The parameters set for alignment: gap opening penalty for pairwise alignment was 35, the gap extension penalty for pairwise alignment was 0.75, the gap opening penalty for multiple alignment was 15, the gap extension penalty for multiple alignment was 0.3, and delay divergent sequences was 25%. After aligning in ClustalX, the alignment output was opened in Bioedit (Tippmann 2004) and adjusted manually. The alignment was refined using both ClustalX and Bioedit.

Phylogenetic Analysis

Unrooted radial Neighbor joining trees are distance based trees constructed from distance matrices of the input multiple alignment file using PAUP4.0b (Hall, 2001). Distances are

calculated from the number of amino acid substitutions that have taken place along a certain branch. PAUP will display the fraction of substitutions between the nodes of that branch as the distances. Bootstrapping was performed by taking subsamples of the sites in the alignment 1000 times and creates trees based on those subsamples (Hall, 2001).

Unrooted radial Parsimony trees (Hall, 2001) were generated using parsimony as the optimality criterion. Parsimony looks for tree or trees with minimum number of changes, which is based on the assumption that is the one requiring the fewest number of changes to explain the data in the alignment. Using random trees as initial trees and 0 as the seed value, heuristic search, which is tree-searching strategy, is performed by tree bisection reconnection (TBR) branch-swapping search, with 1000 random-addition-sequence replicate (Hall, 2001).

Midpoint rooting and outgroup rooting are used to root the neighbor joining and parsimony trees. Midpoint rooting designates a point in the middle of the longest branch between the two most distantly related taxa as the common origin of all sequences. Outgroup rooting is done by selecting one or a group of taxa as an outgroup among those sequences as a root (Hall 2001).

RESULTS AND DISCUSSION

Sequential and Phylogenetic Comparisons of plant SS and bacterial and cyanobacterial GS

Protein sequence alignment analysis was performed on bacterial glycogen synthase (GS) proteins and plant starch synthase (SS) proteins, including multiple examples of the various SS isoforms characterized in green algae and plants, as well as some representative GS sequences of other bacterial species, selected in a blast search using the *Synechocystis* GSI gene as template. Many of the plant SSs, except for the *Arabidopsis* and some *Chlamydomonas* SS isoforms, have been studied at the molecular or biochemical level and have clear annotation. Five translated *Arabidopsis* SS genes, ara S1, ara S2, ara S3, ara S4 and ara S5, and four translated *Chlamydomonas* SS genes, chla bS, chla S1, chla S2, chla S3, are included in the study.

Cao et al. (1999) identified seven conserved sequence motifs according to the functional groups defined by Dayhoff and Orcutt (1979). From protein sequence alignment analysis, we found 11 sequence motifs that are highly conserved among all the bacterial glycogen synthases and plant starch synthases (see appendix I). Among the 11 motifs, one motif, the KTGGL 'look-like motif' has been identified as an ADP/ADP-glucose binding site in *E. coli* GS and implicated in the specific functions of SS and GS (Furukawa et al., 1993). It is found near both the N-terminus and the C-terminus, and the structural similarity between the N-terminal and C-terminal motifs implies similar functions in binding ADP/ADP-glucose (Edwards et al., 1999). The C-terminal motif has an arginine instead of lysine in *E. coli*, but both lysine and arginine residues in this motif can interact with the phosphate moiety of ADP-glucose through ionic interactions. In the N-terminal KTGGL, threonine, which apparently does not interact directly with the ADP/ADP-glucose, can be replaced by other amino acids, e.g., by alanine in SS-like GSIs from the cyanobacteria *Synechocystis*, *Anabaena*, and *Nostoc*, and the bacteria *Magnetococcus*, and the GSs of *Chlamydia* and *Chlamydophila*, valine in SSIII/IV isoforms of green algae and plants, and the GS of cyanobacteria *Synechocystis*, *Nostoc*, *Anabaena*, and *Thermosynechococcus*, *Crocosphaera*, serine in *Deinococcus* GS, and plant SSI. *Magnetococcus* and *E.coli* GS, plant SSII and

gbSS retain threonine. Among those amino acids, the alanine - valine and threonine - serine pairs have similar properties, respectively.

For convenience, we named the SS-like glycogen synthase GSI, and other prokaryotic glycogen synthases GSII. *Nostoc*, *Anabaena*, *Crocosphaera* and *Magnetococcus* each have one GS isoform grouped with the *Synechocystis* GSI, so these four glycogen synthases are called GSIs here. Starting with a larger amount of bacterial GS sequences in the study, most of them are found to be GSII isoforms and fall into one monophyletic group (results not shown here), so we reduce the number of bacterial GS sequences within the GSII clade.

Both neighbor joining and parsimony analysis support almost the same unrooted results (Figs. 1 and 6). Bootstrapping analysis shows the confidence of the unrooted grouping pattern (Fig. 2). Rooting both the neighbor joining and parsimony trees by midpoint rooting (Hall 2001) or outgroup rooting using either bacterial GSIIs or bacterial and cyanobacterial GSIIs as the outgroup (Figs. 3-5 and 7-9) also support the same conclusions.

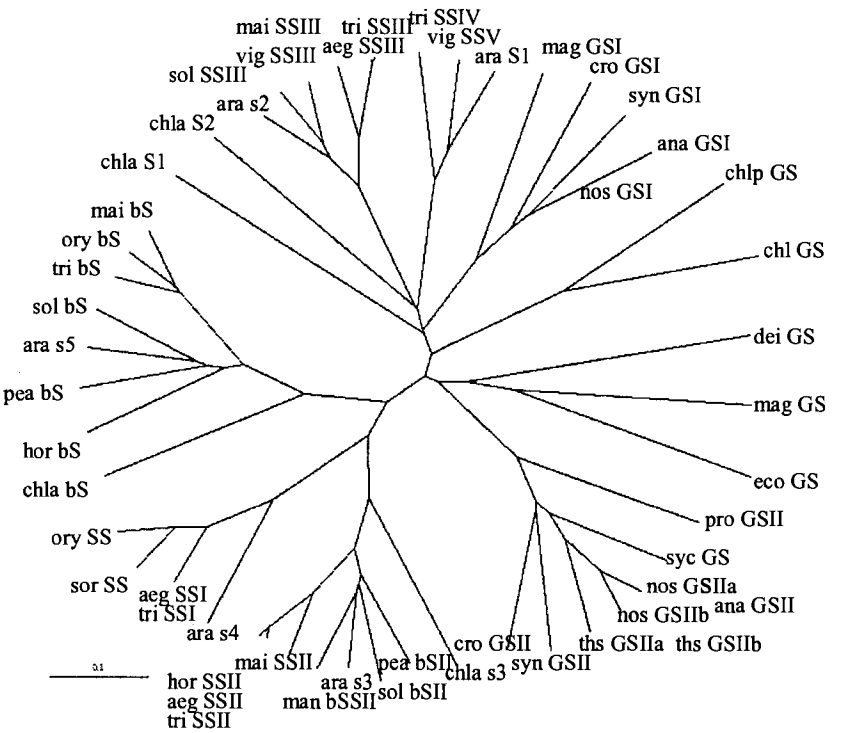

Figure 1 Unrooted Neighbor Joining Tree of Glycogen Synthase

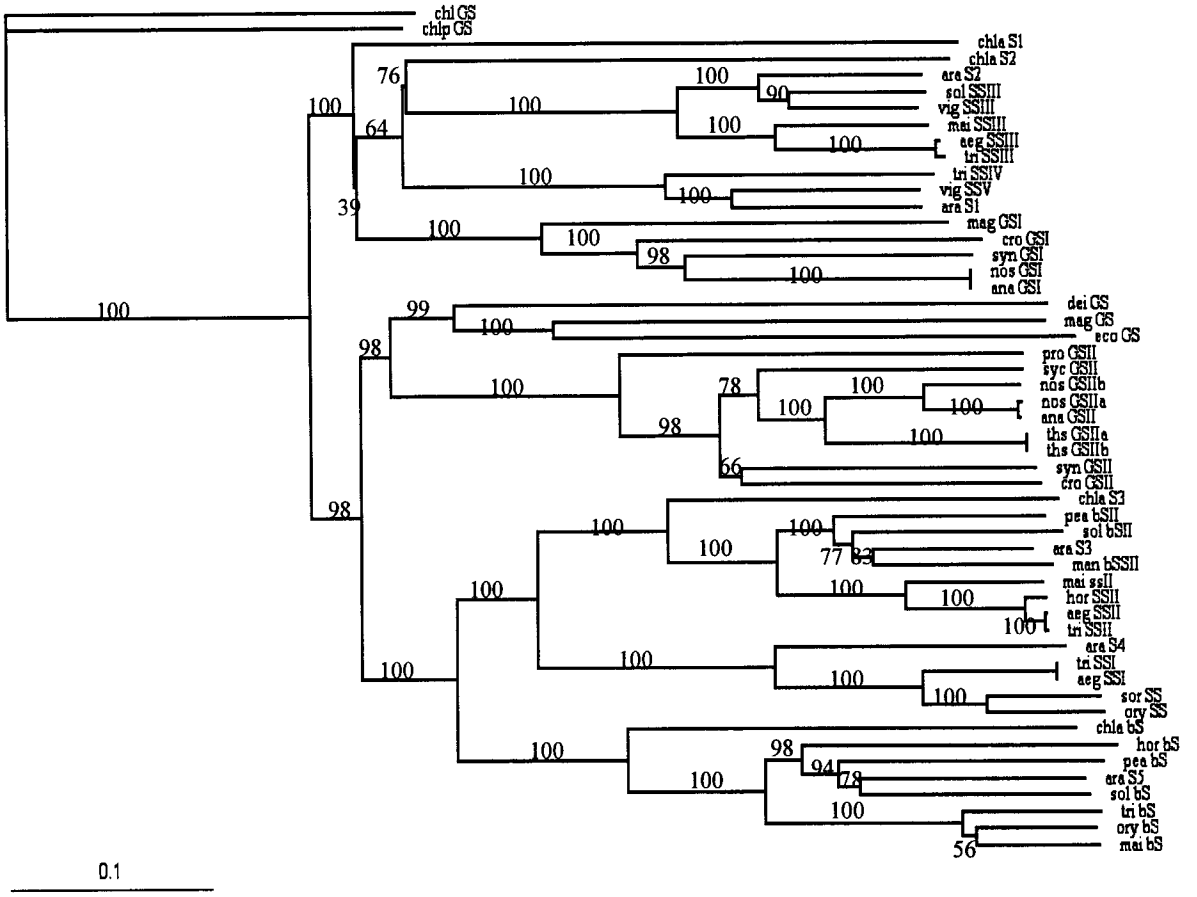

Figure 2 Bootstrap Analysis of Neighbor Joining Tree with Replications of 1000 by Heuristic Search of Glycogen Synthase

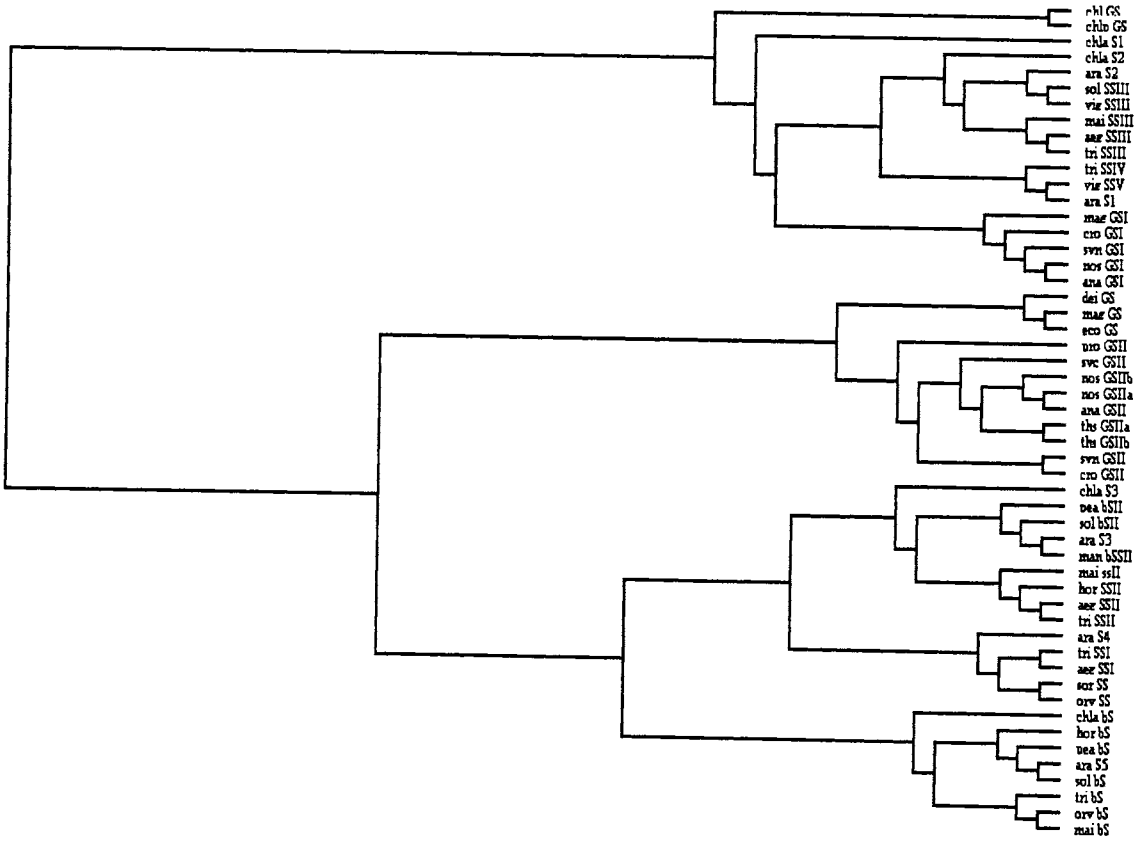

Figure 3 Rooted Neighbor Joining Tree by Midpoint Rooting of Glycogen Synthase

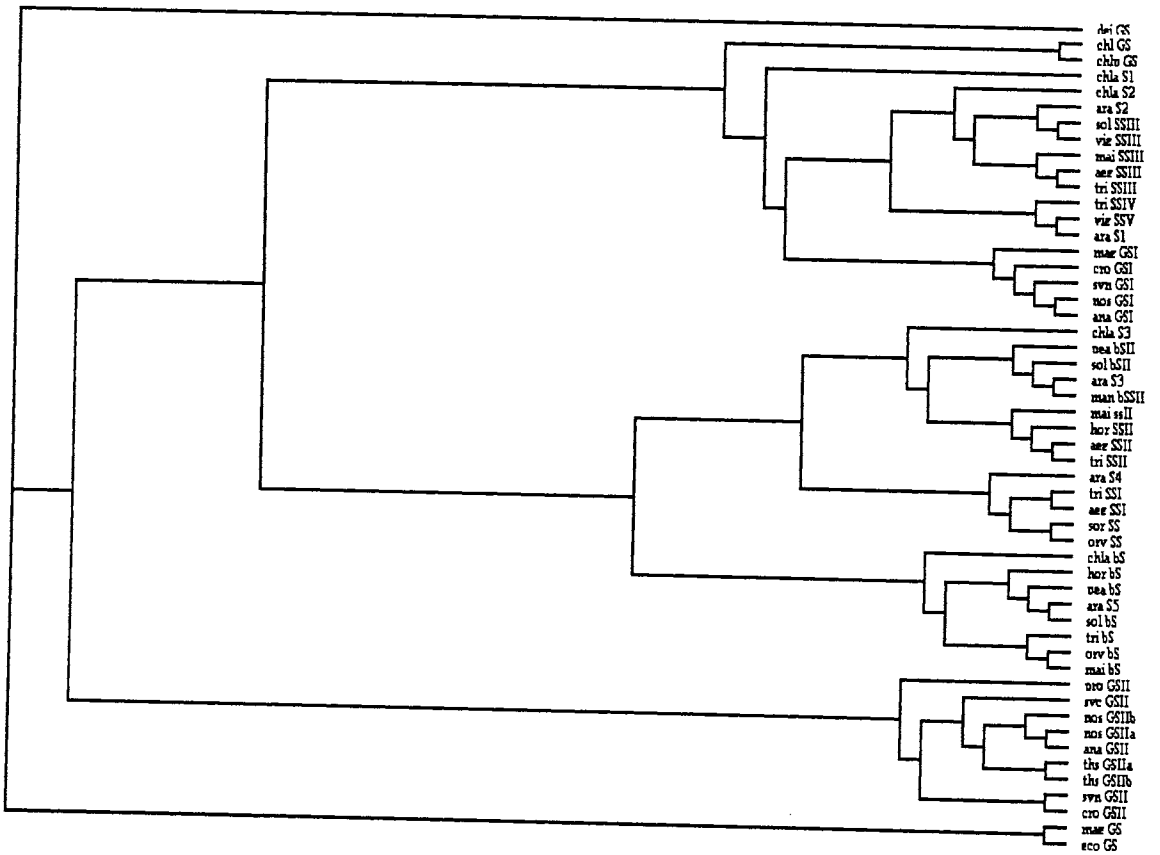

Figure 4 Rooted Neighbor Joining Tree rooting with 'dei GS', 'mag GS', 'eco GS' as outgroup of Glycogen Synthase

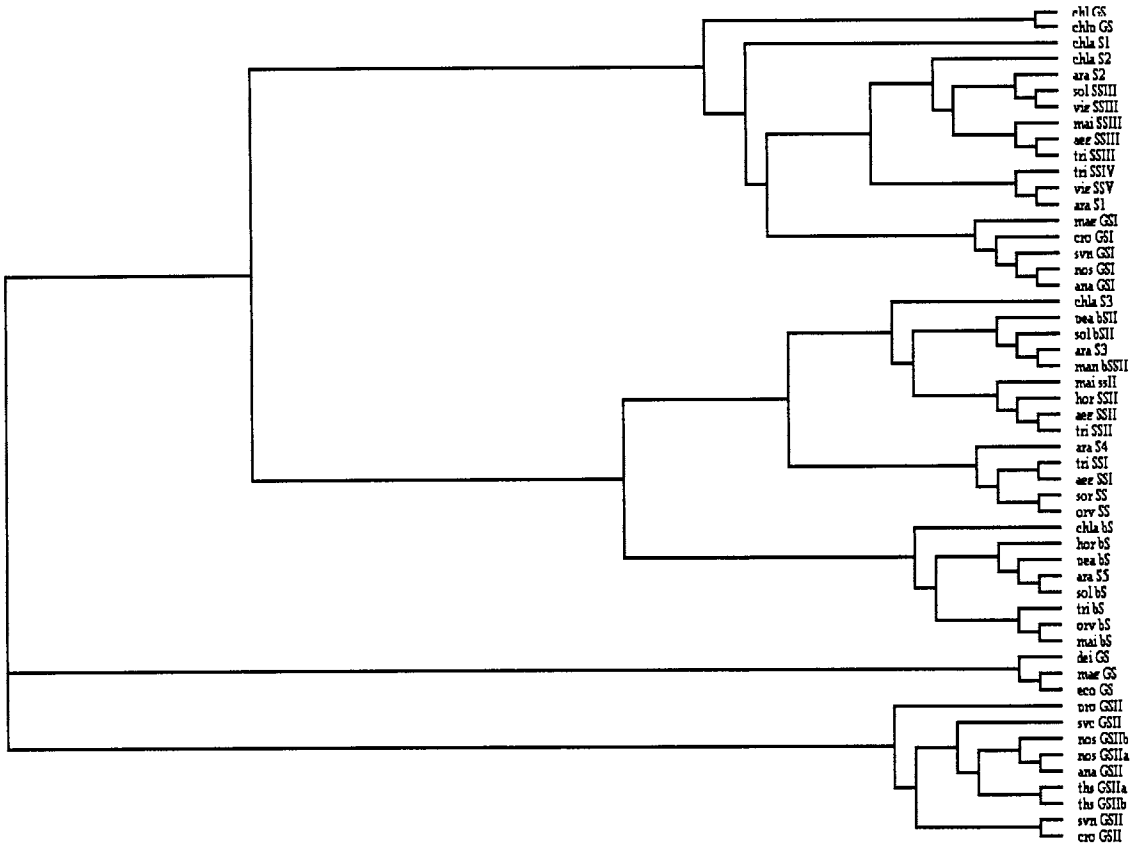

Figure 5 Rooted Neighbor Joining Tree Rooting with 'dei_GS', 'mag_GS', 'eco_GS', 'syn_GSII', 'syc_GSII', 'nos_GSIIa', 'nos_GSIIb', 'ana_GSII', 'ths_GSIIa', 'ths_GSIIb ' as outgroup of Glycogen Synthase

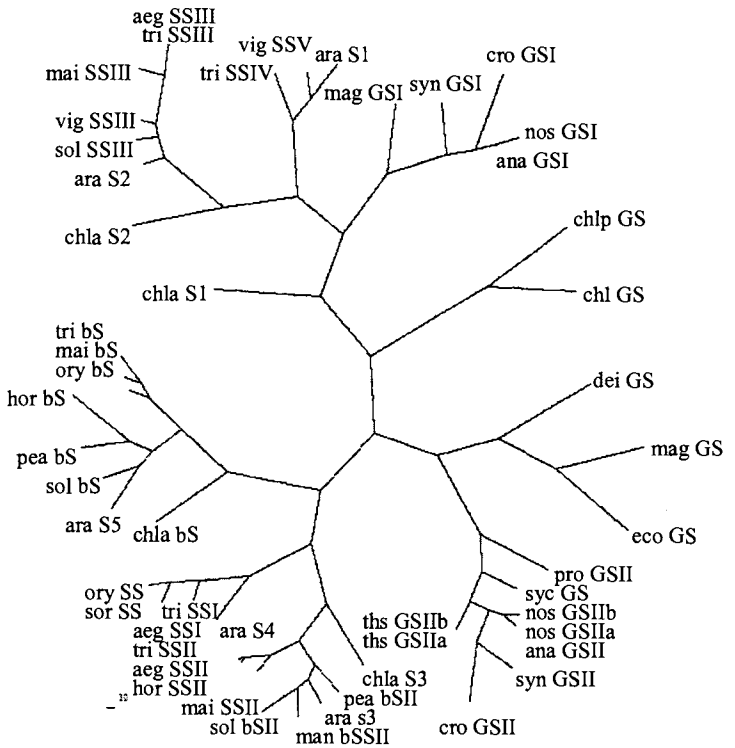

Figure 6 Unrooted Parsimony Tree of Glycogen Synthase

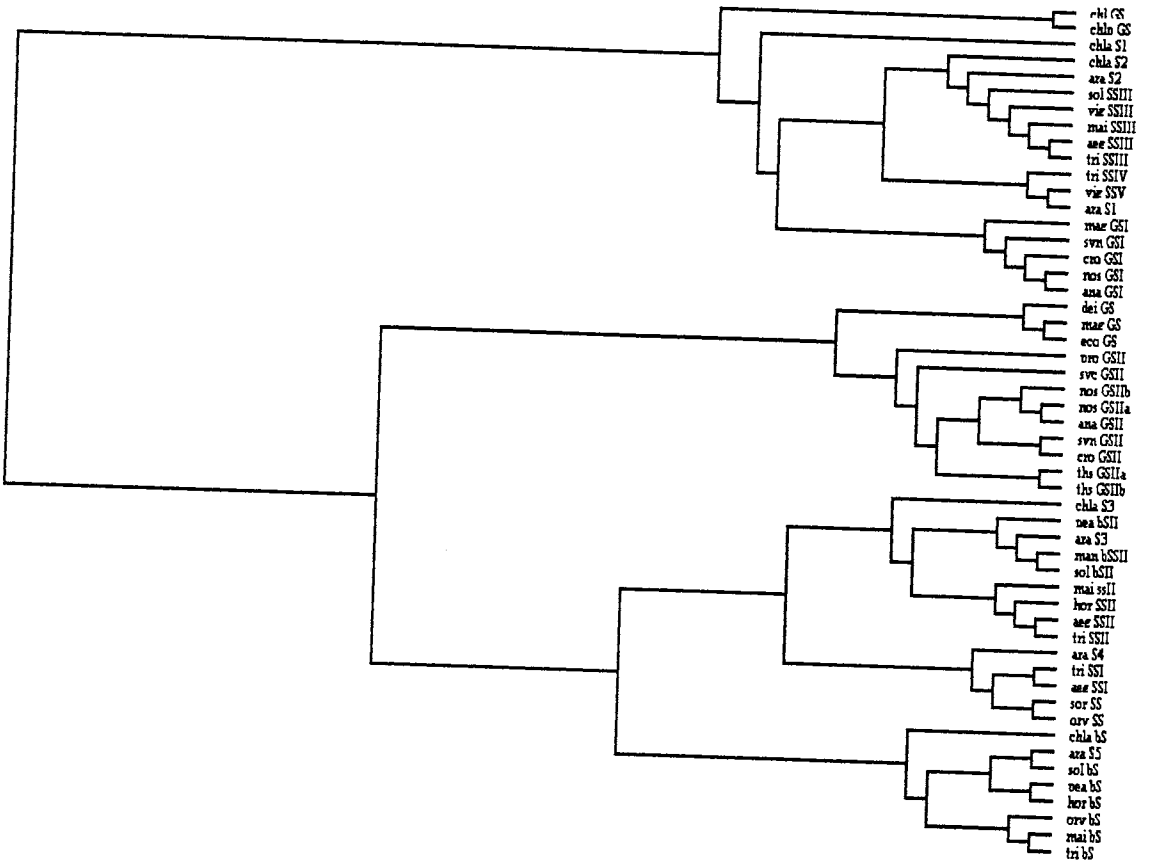

Figure 7 Rooted Parsimony Tree by Midpoint Rooting of Glycogen Synthase

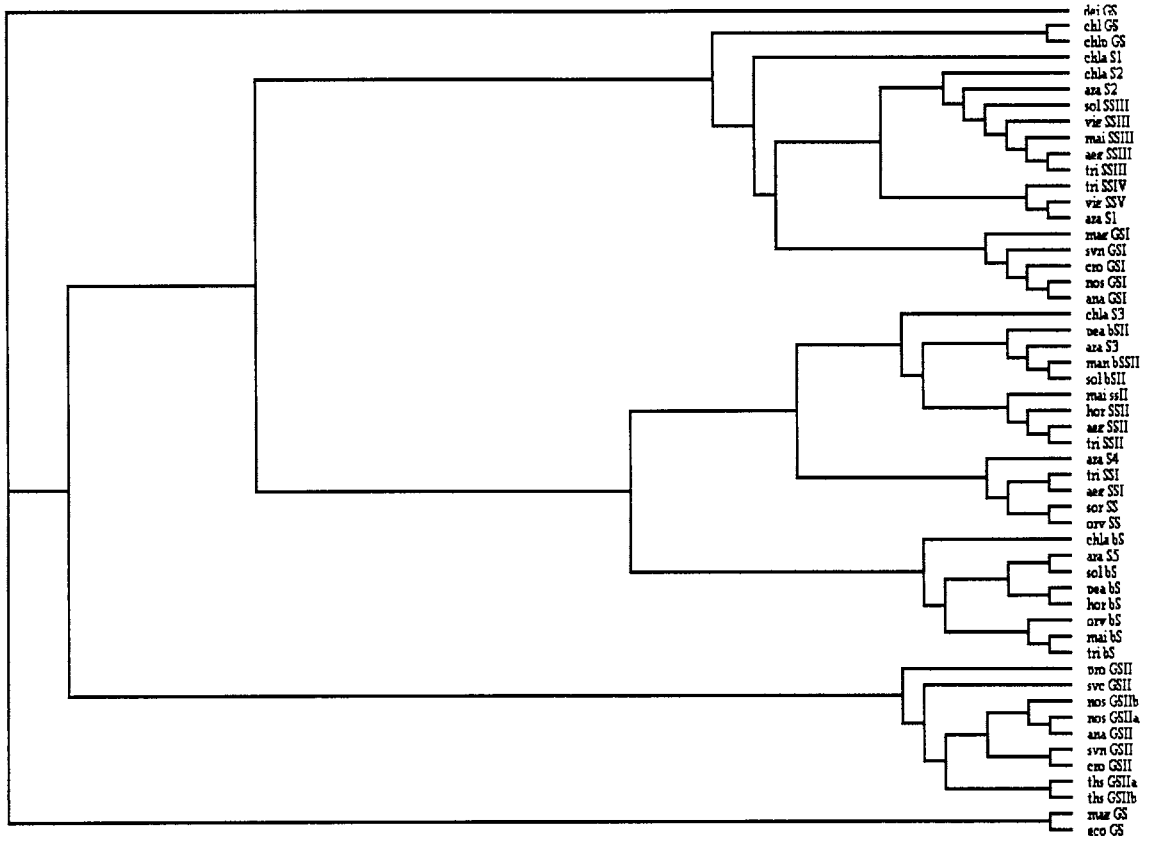

Figure 8 Rooted Parsimony Tree rooting with 'dei_GS', 'mag_GS', 'eco_GS' as outgroup of Glycogen Synthase

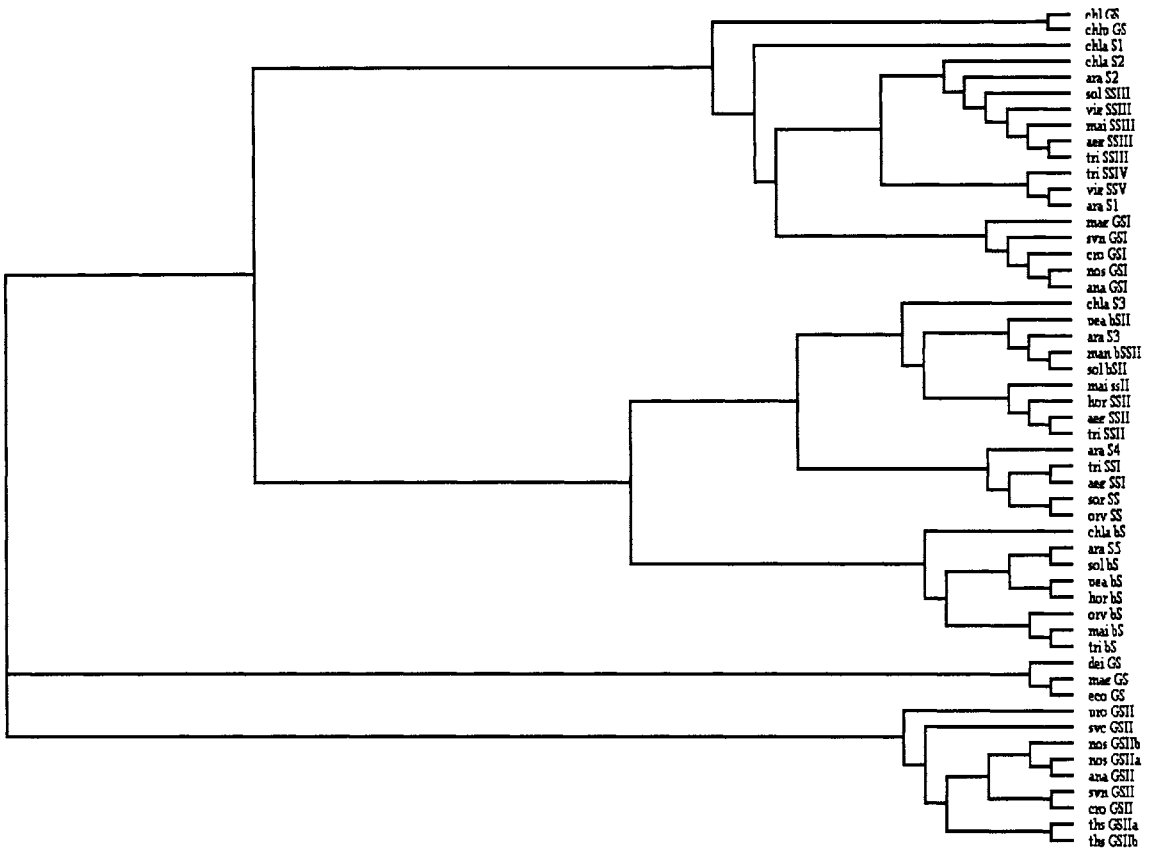

Figure 9 Rooted Parsimony Tree Rooting with 'dei_GS', 'mag_GS', 'eco_GS', 'syn_GSII', 'syc_GSII', 'nos_GSIIa', 'nos_GSIIb', 'ana_GSII', 'ths_GSIIa', 'ths_GSIIb ' as outgroup of Glycogen Synthase

Table 1. Taxon List, Taxonomy and the Accession Number of Sequences of Glycogen Synthase Analysis

Name	Species	Protein	Taxonomy	Accession number
Syn GSI	*Synechocystis*	GSI	Cyanobacteria	gi|16329217|ref|NP_439945.1|
Mag GSI	*Magnetococcus*	GS	Proteobacteria	gi|22999084|ref|ZP_00043088.1|
Vig SSV	*Vigna*	SSV	Eudicots	gi|4582783|emb|CAB40375.1|
Ara S1	*Arabidopsis*	SS	Eudicots	gi|7488349|pir||T04926
Tri SSIV	*Triticum*	SSIV	Monocots	gi|15717885|gb|AAK97773.1|
Chla S1	*Chlamydomonas*	sSS	Green algae	gi|8901183|gb|AAC17971.2|
Ara S2	*Arabidopsis*	SS	Eudicots	gi|15221083|ref|NP_172637.1|
Sol SSIII	*Solanum*	SSIII	Eudicots	gi|2833389|sp|Q43846|UGS4_SOLTU
Vig SSIII	*Vigna*	SSIII	Eudicots	gi|4582789|emb|CAB40374.1|
Mai_SSIII	*Maize*	SSIII	Monocots	gi|7489826|pir||T01265
Aeg SSIII	*Aegilop*	SSIII	Monocots	gi|9502145|gb|AAF88000.1|
Tri SSIII	*Triticum*	SSIII	Monocots	gi|9502143|gb|AAF87999.1|AF258608_1
Chl GS	*Chlamydia*	GS	Chlamydiaceae	gi|15834801|ref|NP_296560.1|
Chlp GS	*Chlamydophila*	GS	Chlamydiaceae	gi|29840578|ref|NP_829684.1|
Mag GS	*Magnetococcus*	GS	Proteobacteria	gi|22998477|ref|ZP_00042543.1|
Pro GS	*Prochlorococcus*	GS	Cyanobacteria	gi|33240502|ref|NP_875444.1|
Eco GS	*E.coli*	GS	Enterobacteria	gi|15803938|ref|NP_289974.1|
Pea bSII	*Pea*	gbSSII	Eudicots	gi|2833384|sp|Q43093|UGS3_PEA
Chla S3	*Chlamydomonas*	sSS	Green algae	gi|8708896|gb|AAC17970.2|
Nos GS2	*Nostoc*	GS	Cyanobacteria	gi|23124653|ref|ZP_00106629.1|
Ara S3	*Arabidopsis*	SS	Eudicots	gi|15232051|ref|NP_186767.1|
Hor SSII	*Hordeum*	SSII	Monocots	gi|23476267|gb|AAN28309.1|
Trit SSI	*Triticum*	SSI	Monocots	gi|9369336|emb|CAB99210.1|
Aeg SSI	*Aegilops*	SSI	Monocots	gi|6103327|gb|AAF03557.1|
Aeg SSII	*Aegilops*	SSII	Monocots	gi|23476265|gb|AAN28308.1|
Mai SSII	*Maize*	SSII	Monocots	gi|7489710|pir||T01208
Tri SSII	*Triticum*	SSIIa	Monocots	gi|7529653|emb|CAB86618.1|
Ara S4	*Arabidopsis*	SS	Eudicots	gi|15237934|ref|NP_197818.1|
Syc GS	*Synechoccocus*	GS	Cyanobacteria	gi|30923330|sp|Q935Y7|GLGA_SYNP7
Ths GS1	*Thermosynechoccocus*	GS	Cyanobacteria	gi|22298305|ref|NP_681552.1|
Man bSII	*Manihot*	gbSSII	Eudicots	gi|6467503|gb|AAF13168.1|AF173900_1
Nos GSIIa	*Nostoc*	GS	Cyanobacteria	gi|17229371|ref|NP_485919.1|
Sor SS	*Sorghum*	sSS	Monocots	gi|12019656|gb|AAD45815.2|
Sol_bSII	*Solanum*	gbSSII	Eudicots	gi|2833390|sp|Q43847|UGS3_SOLTU

Table 1. (continued)

Name	Species	Protein	Taxonomy	Accession number
Ory SS	*Oryza*	sSS	Monocots	gi|2833377|sp|Q40739|UGS2_ORYSA
Dei GS	*Deinococcus*	GS	Eubacteria	gi|15805621|ref|NP_294317.1|
Syn GS	*Synechocystis*	GS	Cyanobacteria	gi|16331219|ref|NP_441947.1|
Chla bS	*Chlamydomonas*	gbSSI	Green algae	gi|16716335|gb|AAC17969.3|
Ara S5	*Arabidopsis*	SS	Eudicots	gi|15223331|ref|NP_174566.1|
Sol bS	*Solanum*	gbSS	Eudicots	gi|602594|emb|CAA58220.1|
Ory bS	*Oryza*	gbSS	Monocots	gi|297422|emb|CAA45472.1|
Pea bS	*Pea*	gbSS	Eudicots	gi|2833383|sp|Q43092|UGST_PEA
Hor bSI	*Hordeum*	gbSSI	Monocots	gi|21667442|gb|AAM74054.1|AF486521_1
Tri bSI	*Triticum*	gbSSI	Monocots	gi|11037536|gb|AAG27624.1|AF286320_1
Mai bS	*Maize*	gbSS	Monocots	gi|136757|sp|P04713|UGST_MAIZE
Ana GS	*Anabaena*	GS	Cyanobacteria	From CyanoBase
Ana GSI	*Anabaena*	GSI	Cyanobacteria	From CyanoBase
Ths GS2	*Thermosynechococcus*	GS	Cyanobacteria	From CyanoBase
Chla S2	*Chlamydomonas*	SS	Green algae	From Chlamydomonas genome site
Cro GSI	*Crocosphaera*	GSI	Cyanobacteria	From JGI
Cro GSII	*Crocosphaera*	GSII	Cyanobacteria	From JGI

The GSI isoforms of *Magnetococcus*, *Synechocystis*, *Anabaena*, *Nostoc* and *Crocosphaera* are in the same clade (SSIII/IV clade) with the SSIII isoforms of *Solanum*, *Aegilop*, *Vigna*, maize, and *Triticum*, the SSIV isoform of *Triticum*, the so-called SSV isoform of *Vigna*, the *Arabidopsis* S2 and SS proteins, and the *Chlamydomonas* S1 and S2 proteinsm with the probability of 100% in bootstrap analysis with replications of 1000. *Chlamydia* GS and *Chlamydophila* GS form a sister clade to the SSIII/IV clade. *Deinococcus*, *Magnetococcus* and *E. coli*, which are three representative kinds of bacteria, form another, separate clade (bacterial GSII clade). The GSII proteins from the cyanobacteria *Prochlorococcus*, *Synechocystis*, *Synechococcus*, *Thermosynechococcus*, *Nostoc* and *Anabaena* comprise another clade (cyanobacterial GSII clade). The gbSS isoforms of monocots, such as maize, *Oryza*, *Triticum*, dicots, such as *Solanum*, *Arabidopsis*, pea, *Hordeum*, and the green alga, *Chlamydomonas*, are grouped into a clade, although there is a clear divergence between the monocots and dicots in this clade. SSI and SSII form sister clades containing the *Oryza*, *Sorghum*, and *Aegilop* SSI isoforms, along with *Arabidopsis* S4 in the SSI clade, and the maize, *Aegilop*, and *Manihot* SSII isoforms, and the so-called bSSII isoforms of *Solanum* and

Manihot, along with *Chlamydomonas* S3 in the SSII clade. Amino acids predicted for each taxa to have changed, according to the character change analysis (Maddison. 1989), from their nearest ancestors within the GSI and SSIII/IV clade are listed in Table 2. Those changes provide supportive evidence that GSI and SSIII/IV should be grouped into one clade and do indeed share a close evolutionary relationship.

The same SS or GS isoforms, even though from different species appear in the same clade, which indicates that gene duplication events leading to development of these different isoforms occurred before speciation. According to the grouping pattern, the *Arabidopsis* S2 should be an SSIII isoform, the *Arabidopsis* S1 should be a SSIV isoform, *Manihot* and *Solanum* bSSII, and *Arabidopsi*s S3 should be SSII isoforms, *Arabidopsi*s S4 should be an SSI isoform, and *Arabidopsi*s S5 should be a granule-bound SS isorforms. *Chlamydomonas* S1 and S2 appear to be either SSIII, or SSIV isoforms.

Since isoforms SSIII and SSIV show a closer evolutionary relationship with GSI than with other SS isoforms, they may share a more recent common ancestor with these cyanobacterial GSI genes. Bacterial GSIIs or some ancestral GS genes may have given rise to GSI isoforms or GSI ancestral genes via gene duplication. Because the GSI isoforms have higher similarity than GSII isoforms or than SSI, SSII or gbSS isoforms, to the SSIII and SSIV isoforms of photosynthetic eukaryotes that would appear later in evolution, the SSIII/SSIV group, at least, probably arose from an ancestral GSI-like gene.

Since the SSIII and SSIV isoforms are not found in prokaryotes, this gene duplication must have occurred subsequent to the endosymbiotic event leading to the formation of plastids. After the emergence of the GSI ancestral genes, the entry of cyanobacteria into eukaryotic cells through endosymbiosis gave rise to the ancestor of plastids. Since GSI isoforms are quite similar to some isoforms of starch synthases, an ancestral GSI isoform might have possessed characteristics beneficial for the emergence of starch biosynthesis. It appears that all photosynthetic eukaryotes have both SSIII and SSIV isoforms, so duplication of the GSI ancestral gene must have occurred early in the development of plastids to produce ancestral forms of SSIII and SSIV. By gene trafficking, either the ancestral GSI gene or the ancestral SSIII and SSIV genes were transferred to the nucleus to become nuclear-encoded genes, as

SSIII and SSIV are today. Transfer to the nucleus could have occurred either before or after gene duplication events that gave rise to SSIII and SSIV.

SSI, SSII and gbSS isoforms occur in a clade unrelated to the SSIII/IV clade in the phylogenetic analysis reported here, thus their origin must be different. These isoforms may have arisen from duplication of the ancestral GS genes of the eukaryotic host cell, which may already have been present in the nucleus, or they may have arisen from an ancestral GSII-like gene also likely to have been present in the endosymbiont.

Table 2. Amino Acids Change of GSI and SSIII/IV Clade from Its Nearest Ancestor of Glycogen Synthase Analysis. Those Amino acids Shown in Bold Lie within the 11 Conserved Regions Identified.

Position	Ancestoral aa	Changed aa
2	**N**	**H**
6	**V**	**I**
10	**C**	**M**
24	**G**	**T**
46	K	C
50	S	D
51	L	Q
55,56	AD	LK
60	K	D
67	M	P
85	Y	V
88	E	G
95	Y	V
98	D	I
100	H	P
103	F	D
109	N	F
121	G	Y
129	**L**	**E**
180	Y	V
188,189	RY	SR
208	E	D
237	D	A
266	**D**	**N**
279	**I**	**V**
291	L	T
296	S	E
322	A	P

Table 2. (continued)

Position	Ancestor aa	Changed aa
394	**E**	**P**
399	**D**	**H**
408	**L**	**T**
416	V	G
437	L	A
471	**M**	**V**
483	**L**	**M**
505	K	F
507	G	V
521	G	R
528	R	G
572	T	L
578	P	E
609	L	I

In the SSII, gbSS, and SSIII/IV clades, *Chlamydomonas* SSs all diverge earlier than plant SSs, which suggests an earlier emergence of *Chlamydomonas* SSs than plant SSs. In fact, *Chlamydomonas* S1, although part of the SSIII/IV clade, emerged too early to clearly be SSIII or SSIV, and *Chlamydomonas* S2 appears perhaps to have at least some affinity to both the SSIII and SSIV isoforms, suggesting that an ancestral *Chlamydomonas* S2-like gene may have been the progenitor duplicated to form the separate SSIII and SSIV isoforms. There is an additional *Chlamydomonas* SS gene, apparently an SSI isoform, identified in the *Chlamydomonas* genome (Shrager et al. 2003), but its sequence was too incomplete to use in this phylogeny. Its presence does indicate though, that even in photosynthetic eukaryotes that emerged earlier than the plants, five SS isoforms already were present.

According to the high degree of morphological differentiation in their multicellular tissue, a five subsection higher-level classification of cyanobacteria has been defined (Castenholz and Waterbury 1989a, 1989b). *Prochlorococcus*, *Synechocystis*, and *Synechococcus* belong to subsection I, unicellular cyanobacteria producing by binary fission or by budding. *Trichodesmium* belongs to subsection III, filamentous nonheterocystous cyanobacteria dividing in only one plane. *Nostoc* and *Anabaena* belong to subsection IV, filamentous heterocystous cyanobacteria dividing also dividing in only one plane (Honda et al., 1999; The Tree of Life Web Project).

The presence of GSI isoform genes only in four of the cyanobacterial genomes thus far sequenced could be explained if the duplication giving rise to GSI occurred in the cyanobacteria group most closely related to the original endosymbiont "pre-plastids". *Synechocystis* is unicellular cyanobacteria and *Nostoc* and *Anabaena* is filamentous cyanobacteria. Duplications may occur in these two subsections of cyanobacteria. However, this would not explain the presence of GSI-like isoforms in *Magnetococcus* and in the *Chlamydiaceae*. There are at least two possible explanations for this apparently scattered presence of GSI isoforms among the bacteria. One possibility is that the gene duplication events that gave rise to GSI or GSI ancestral genes from an ancestral GSII gene only occurred in cyanobacteria and that the GSI-like gene in *Magnetococcus* was acquired through horizontal gene transfer. Bacteria and archaea are known to have the ability to adapt to new environments by the acquisition of new genes through horizontal transfer (Philippe et al., 2003), so the presence of a GSI-like protein in this single, isolated bacterium argues for horizontal gene transfer. Another possibility is that the GSII to GSI duplication occurred early in bacterial evolution and GSI genes in most of the bacteria or cyanobacteria have been lost, except in some species, such as *Magnetococcus*, *Synechocystis*, *Nostoc*, *Anabaena*, *Crocosphaera* and the two bacterial species in the Chlamydiaceae. The presence of GSI genes in the Chlamydiaceae will be more fully discussed below.

Whether the duplication of GSIs occurred early in bacterial evolution or only in the cyanobacteria closely related to the original endosymbionts is still a puzzle to be solved, but, based on the assumption that the most likely evolutionary process is the one that requires the fewest number of duplications or deletions, duplication of an ancestral GSII-like gene to form an ancestral GSI-like gene only in the cyanobacteria seems most likely. However, only a relatively small number of bacterial and cyanobacterial genomes have been sequenced so far, so we do not know how many bacterial species other than *Magnetococcus* also have GSI isoforms. With more bacterial genomes being sequenced in the future, this sample size will continue to grow and with it the answer to this question may emerge.

Sequence and Phylogenetic Analysis of Isoamylase-type Glycogen Debranching Enzymes

This phylogenetic analysis of plant IA and bacterial IA-type GDBE included multiple examples of the three IA isoforms identified in plants, as well as a large number of bacterial GS sequences, most of which were selected in a blast search using the *Synechocystis* type I GDBE (syn1) as template. Thus all the bacterial genes represent IA-related GDBEs, rather than the whole spectrum of bacterial GDBEs, including pullulanase-type GDBEs and perhaps other types. Because the plant IA isoforms are the SDBE types most clearly implicated in starch synthesis (Hussain et al., 2003), as opposed to only hydrolysis and catabolism, only the IA-like proteins from bacteria have been included. Sequence analysis shows 13 conserved motifs in those sequences analyzed among species from bacteria to higher plants (see appdendix II), such as GIEVIL, but there are no published studies on conserved motifs in IAs yet.

Both unrooted neighbor joining and unrooted parsimony analysis support almost the same results (Figs. 10 and 16). Bootstrapping analysis shows the confidence of the grouping pattern (Fig. 11). Rooting both the neighbor joining and parsimony trees by midpoint rooting or by outgroup rooting using either bacterial GDBEs or bacterial and cyanobacterial GDBEs as outgroup (Figs. 12-15 and 17-20) also support the same basic conclusions.

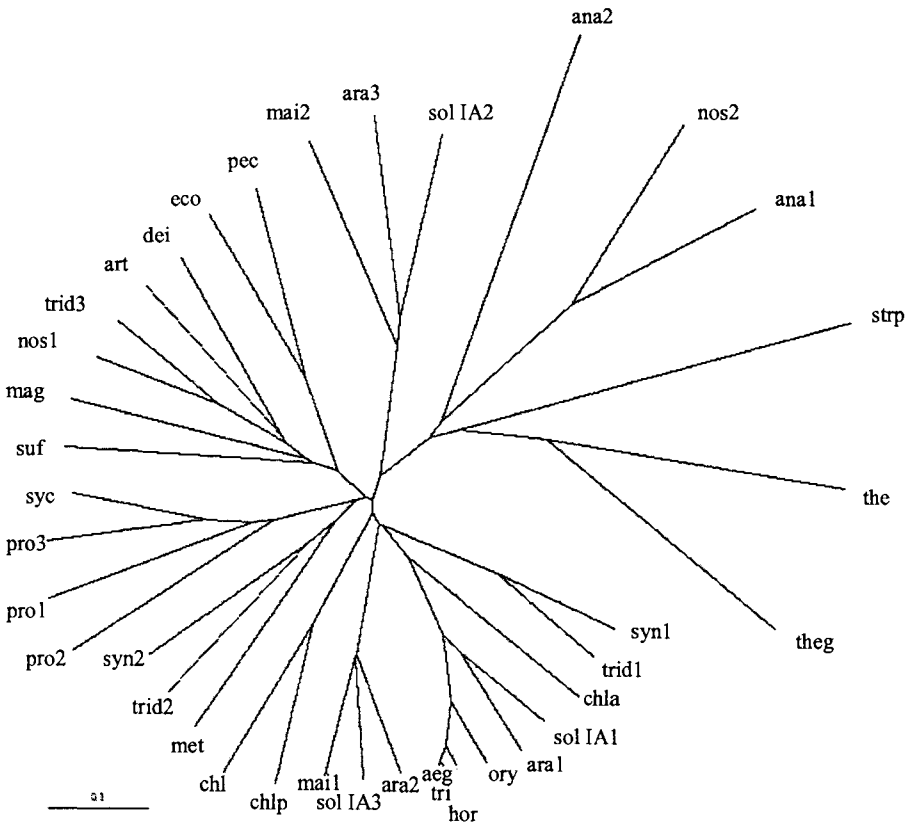

Figure 10 Unrooted Neighbor Joining Tree of Isoamylase-type Debranching Enzyme

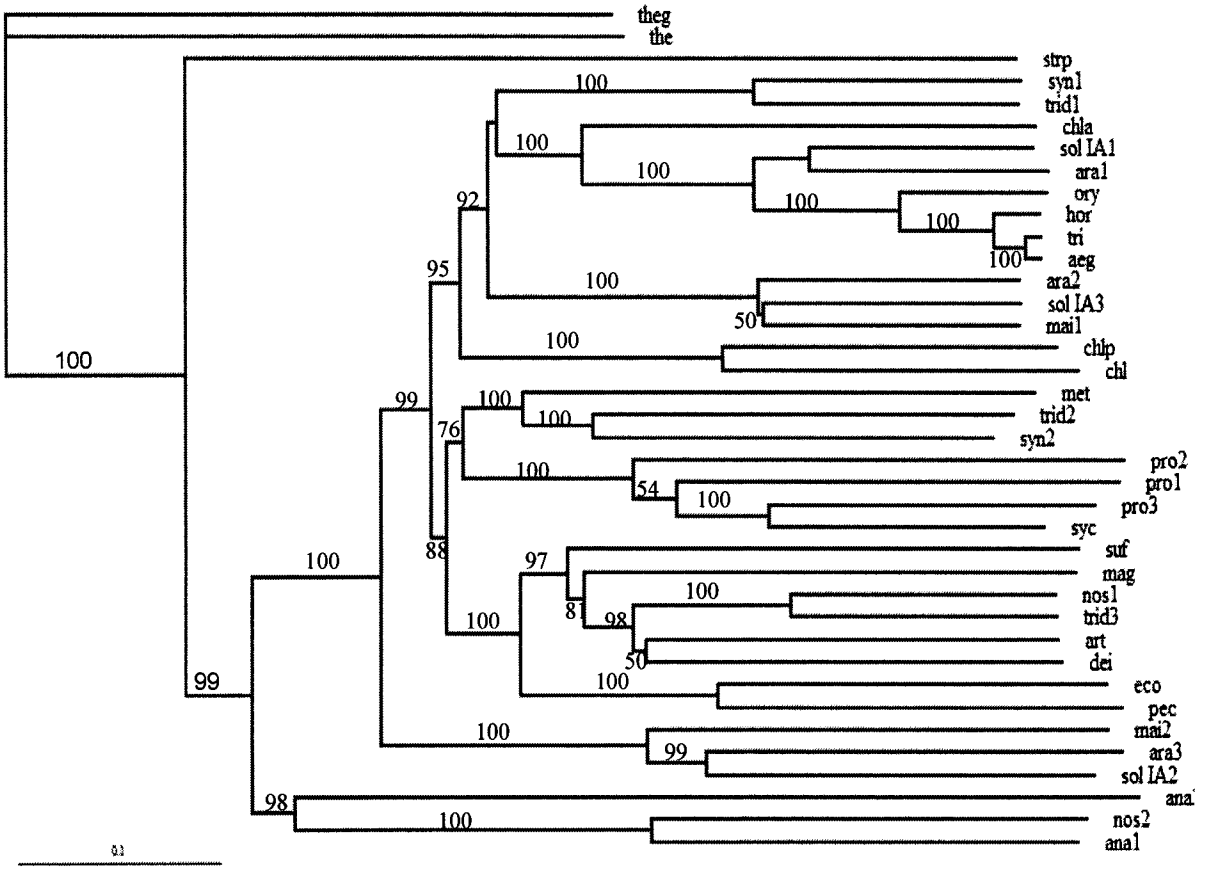

Figure 11 Bootstrap Analysis of Neighbor Joining Tree with Replications of 1000 by Heuristic Search of Isoamylase-type Debranching Enzyme

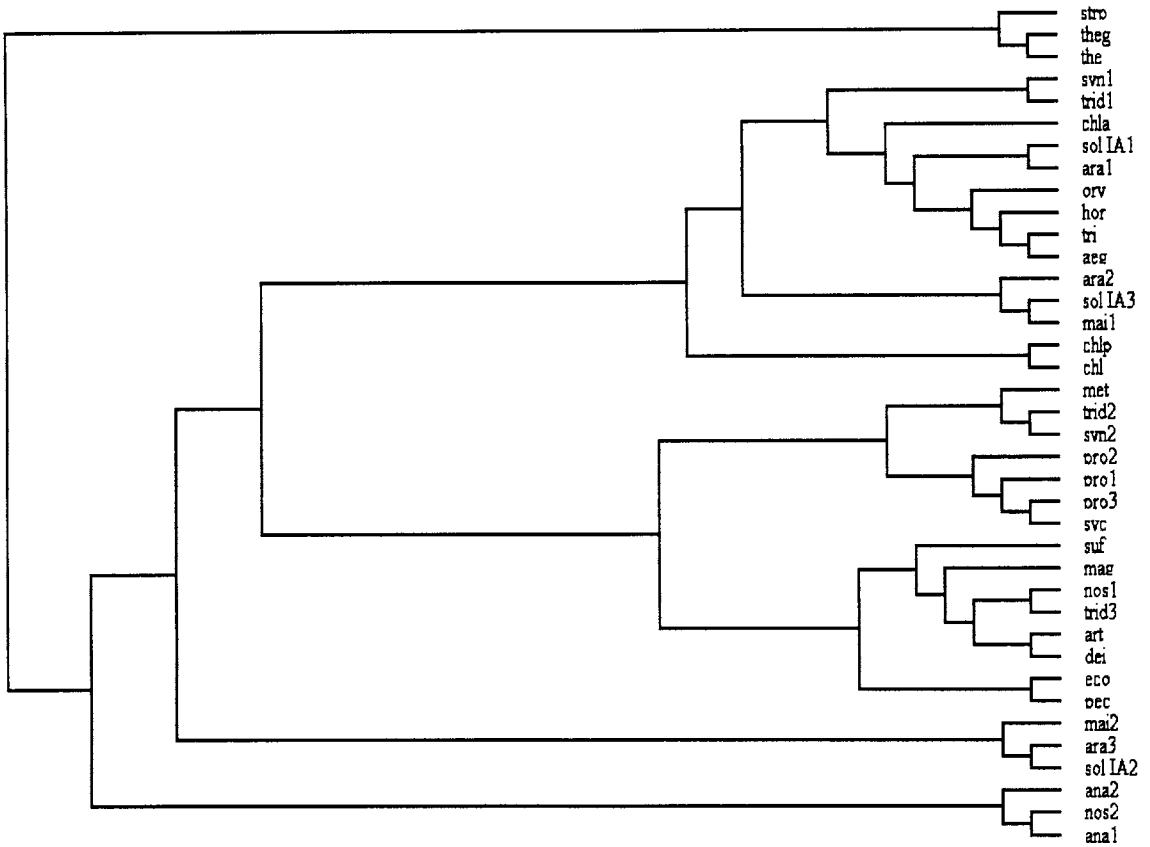

Figure 12 Rooted Neighbor Joining Tree by Midpoint Rooting of Isoamylase-type Debranching Enzyme

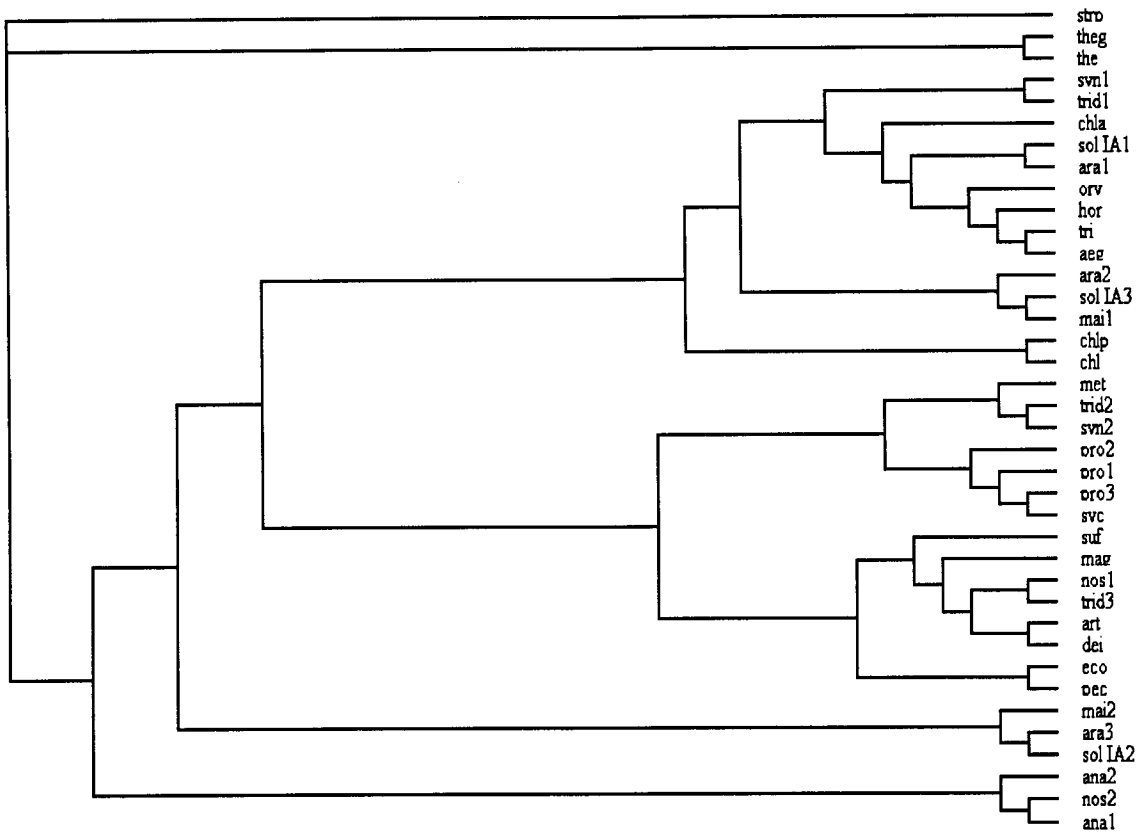

Figure 13 Rooted Neighbor Joining Tree rooting with 'strp', 'theg', 'the' as outgroup of Isoamylase-type Debranching Enzyme

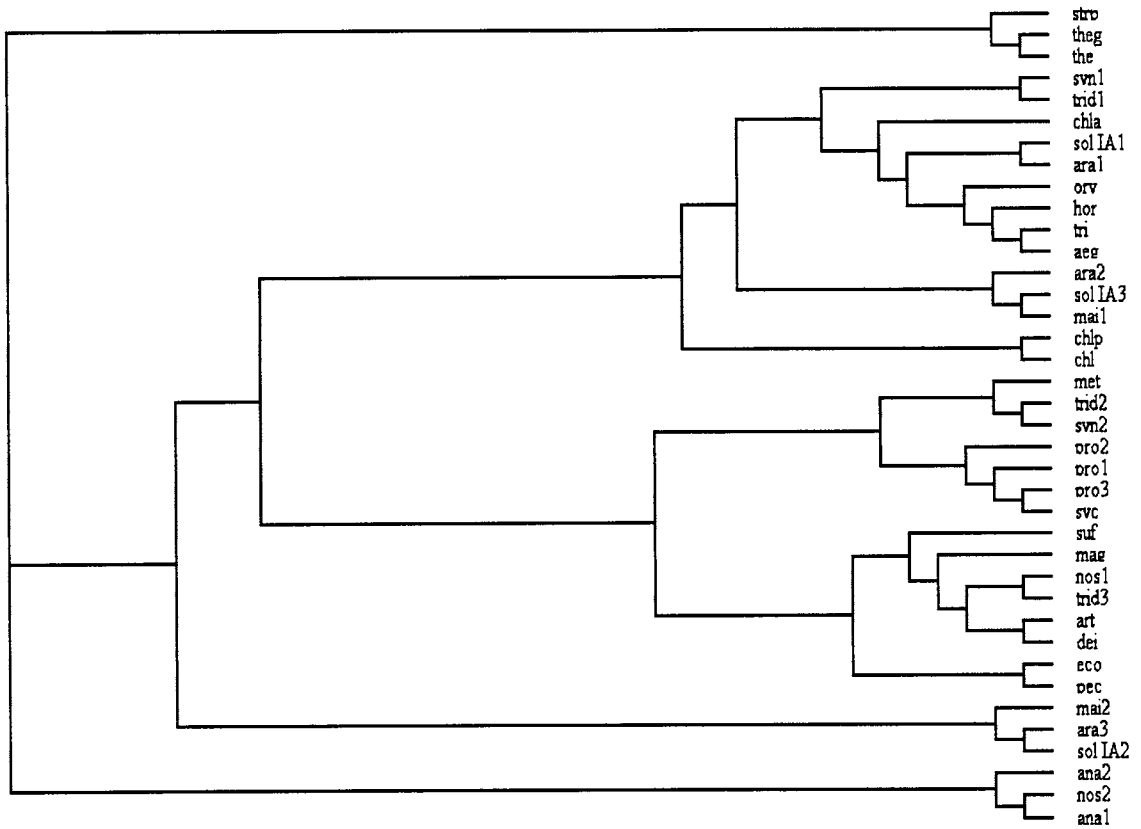

Figure 14 Rooted Neighbor Joining Tree rooting with 'strp', 'theg', 'the', 'ana2', 'nos2', 'ana1' as outgroup of Isoamylase-type Debranching Enzyme

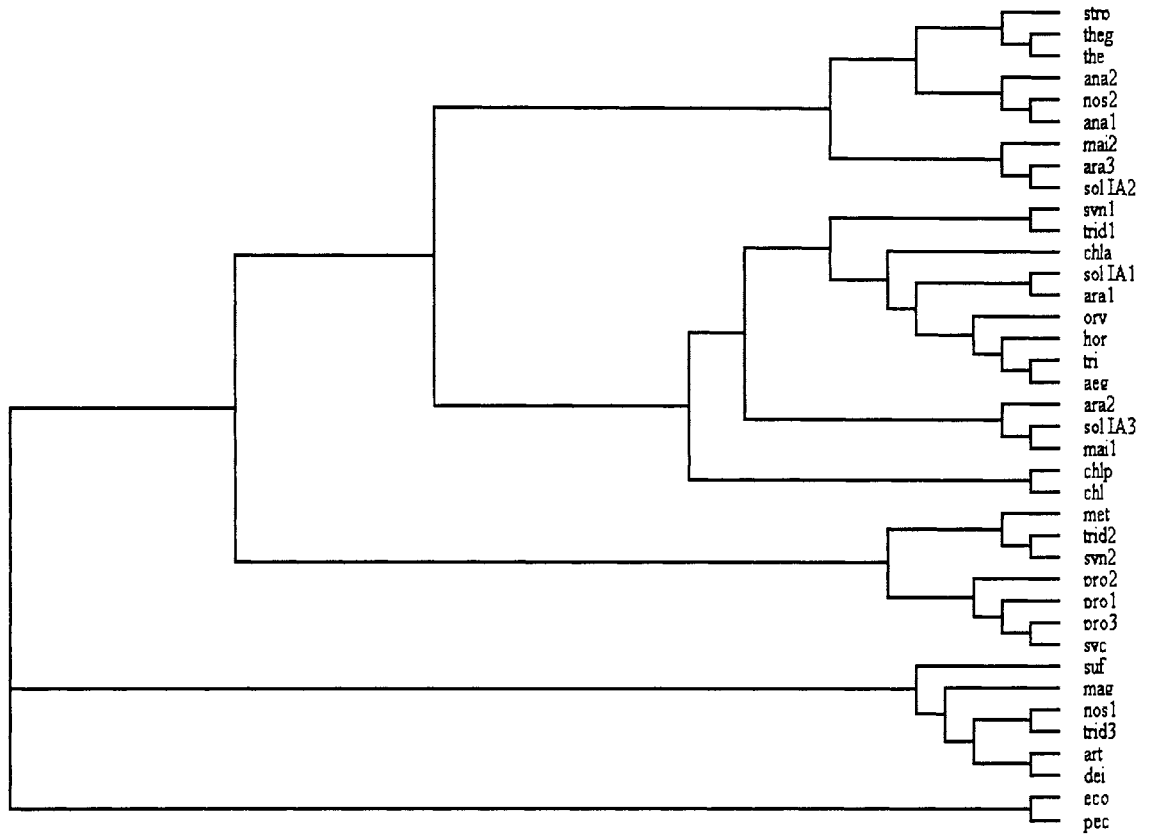

Figure 15 Rooted Neighbor Joining Tree rooting with 'suf', 'mag', 'nos1', 'trid3', 'art', 'dei', 'eco', 'pec' as outgroup of Isoamylase-type Debranching Enzyme

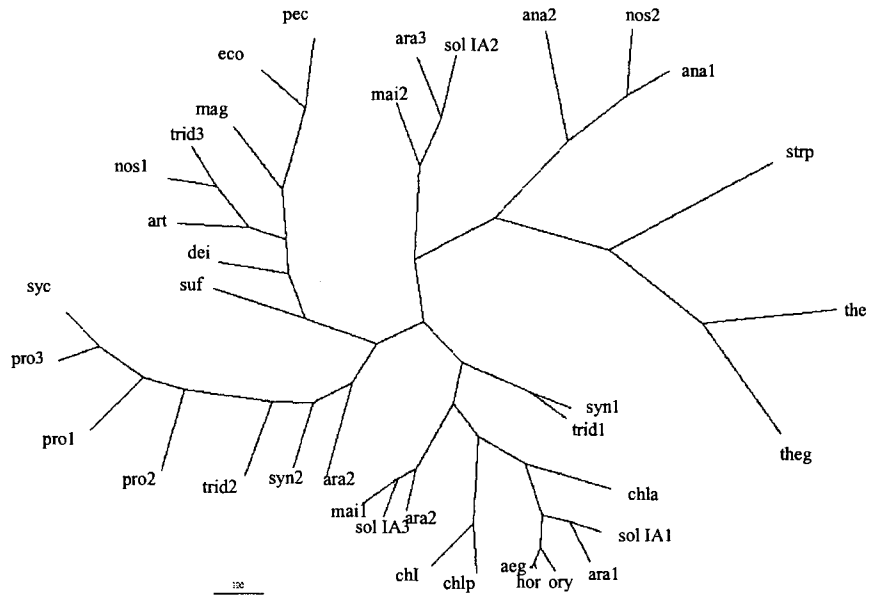

Figure 16 Unrooted Parsimony Tree based on Parsimony Analysis on Isoamylase-type Debranching Enzyme of Isoamylase-type Debranching Enzyme

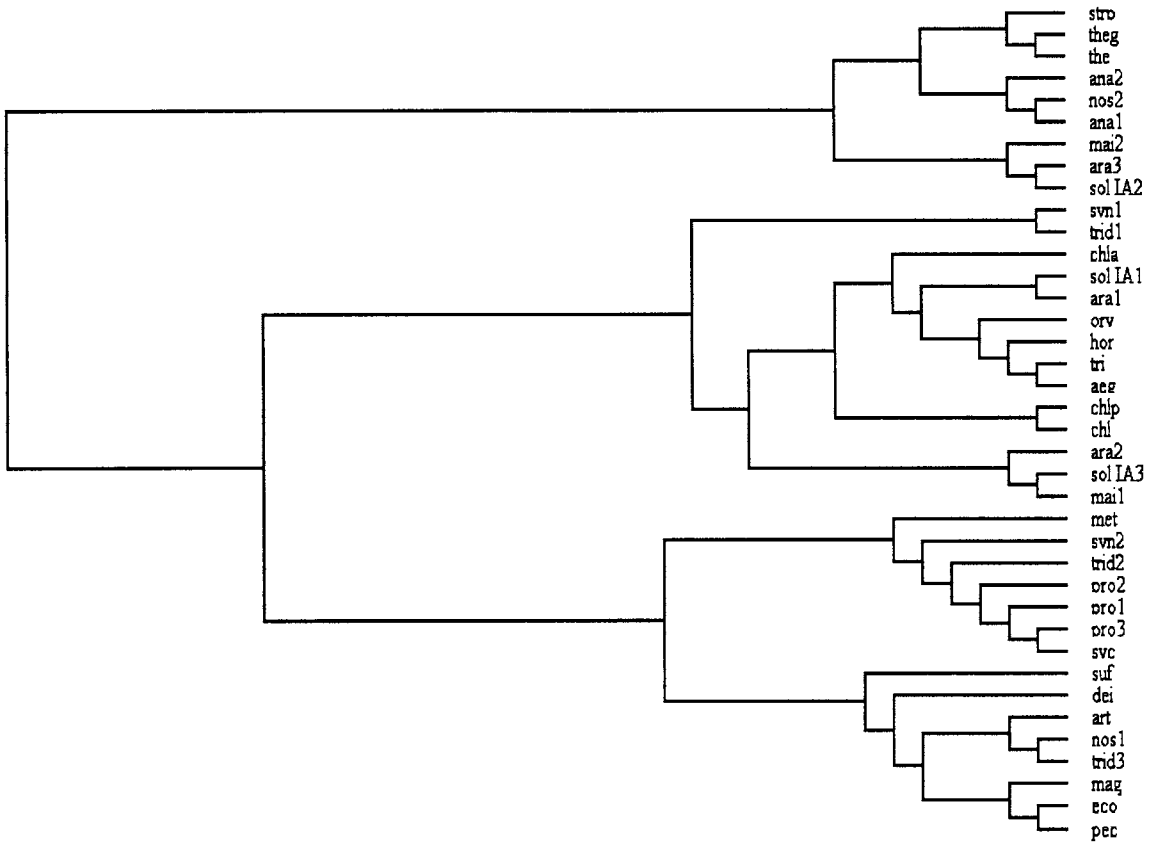

Figure 17 Rooted Parsimony Tree by Midpoint Rooting of Isoamylase-type Debranching Enzyme

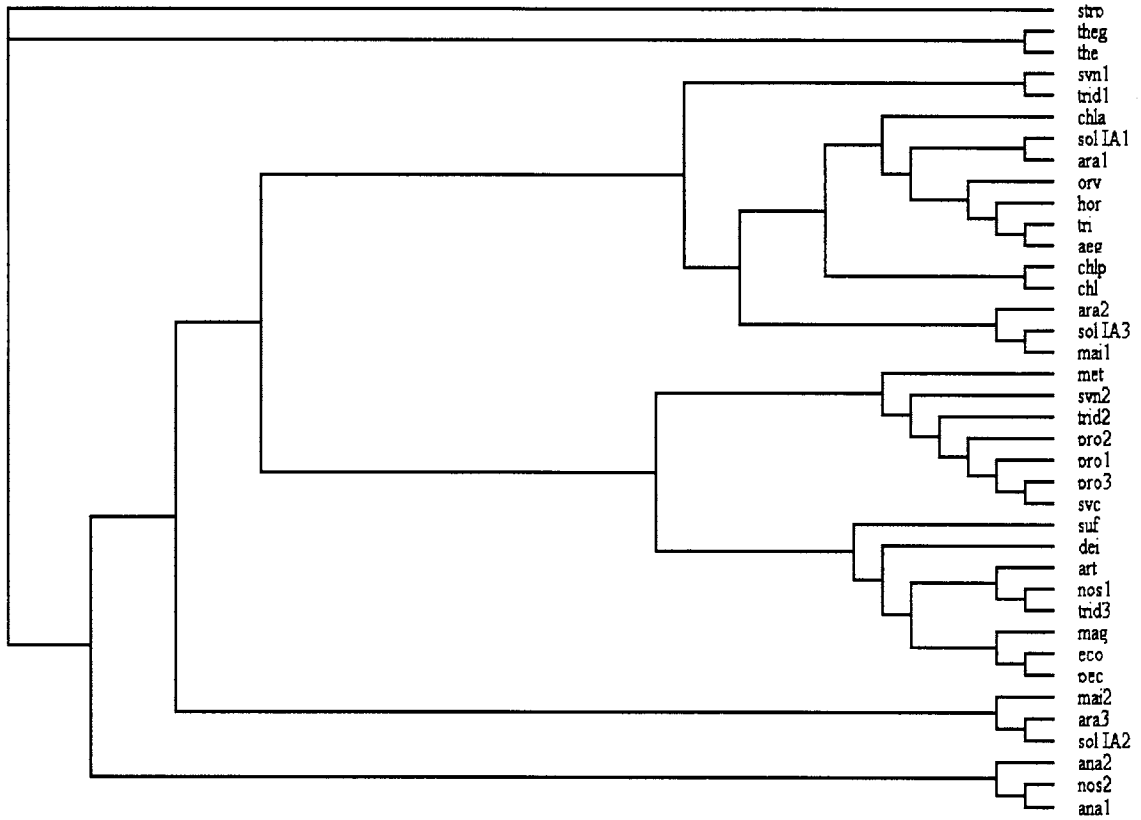

Figure 18 Rooted Parsimony Tree Rooting with 'strp', 'theg', 'the' as outgroup of Isoamylase-type Debranching Enzyme

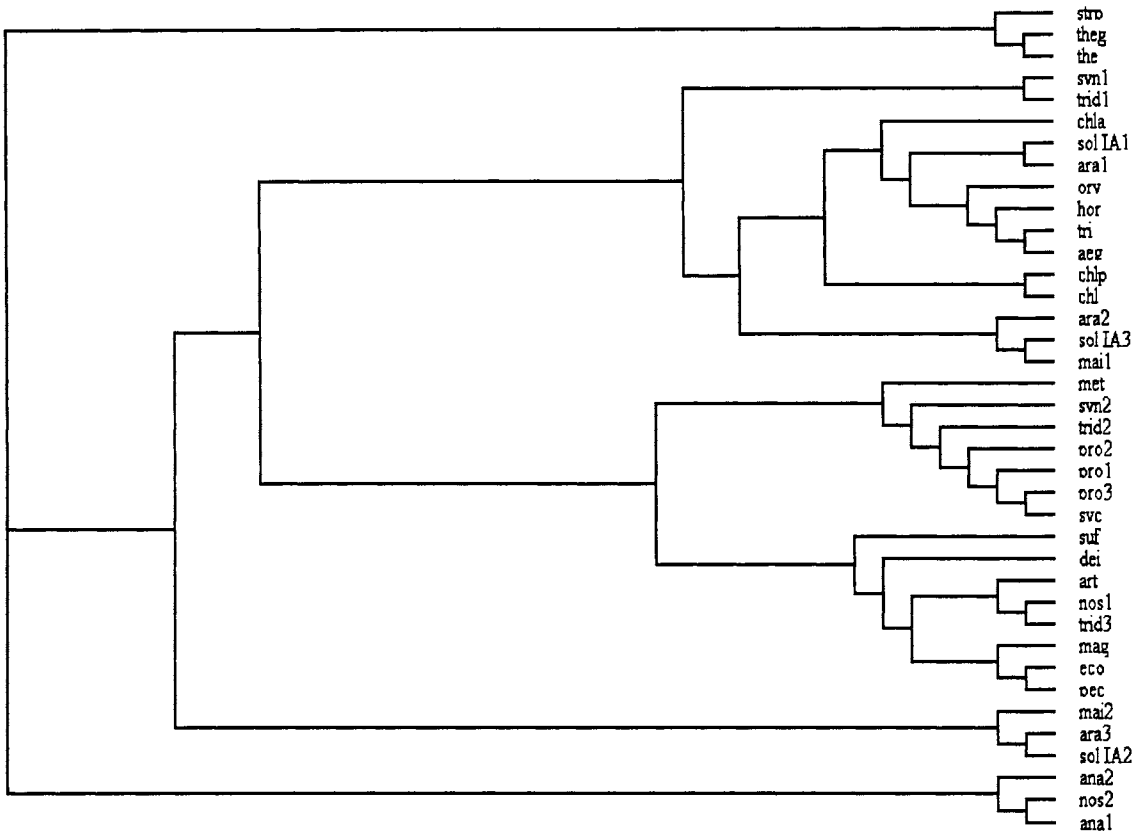

Figure 19 Rooted Parsimony Tree rooting with 'strp', 'the', 'the', 'ana2', 'nos2', 'ana1' as outgroup of Isoamylase-type Debranching Enzyme

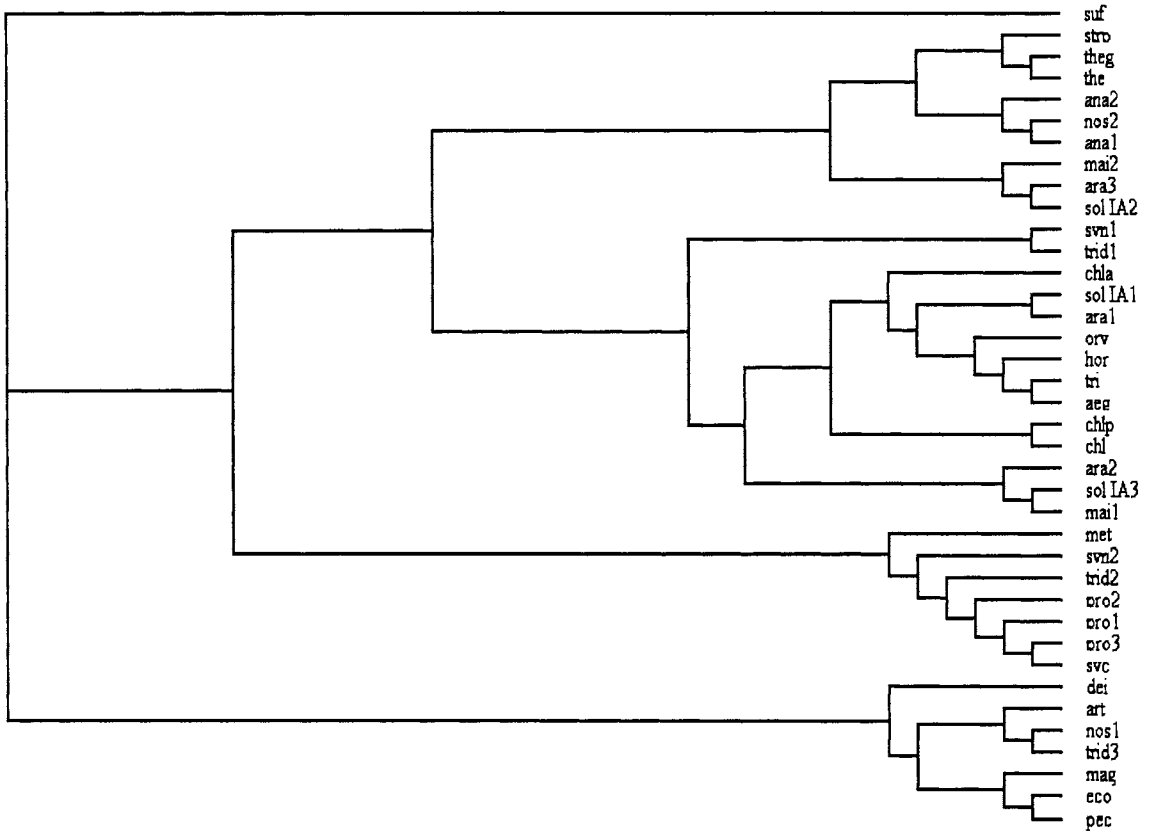

Figure 20 Rooted Parsimony Tree rooting with 'suf', 'dei', 'art', 'mag', 'nos1', 'trid3', 'eco', 'pec' as outgroup of Isoamylase-type Debranching Enzyme

Table 3. Taxon List, Taxonomy and the Accession Number of Sequences of Isoamylase-type Debranching Enzyme Analysis

Name	Species	Protein	Taxonomy	Accession number
Syn1	Synechocystis		Cyanobacteria	gi|16330244|ref|NP_440972.1|
Trid1	Trichodesmium		Cyanobacteria	gi|23042506|ref|ZP_00073886.1|
Sol_IA1	Solanum	IA1	Eudicots	gi|27728145|gb|AAN15317.1|
Tri	Triticum		Monocots	gi|17932898|emb|CAC82925.1|
Hor	Hordeum		Monocots	gi|21314275|gb|AAM46866.1|
Ara1	Arabidopsis		Eudicots	gi|15225595|ref|NP_181522.1|
Aeg	Aegilops		Monocots	gi|31096626|gb|AAP44579.1|
Trid2	Trichodesmium		Cyanobacteria	gi|23043137|ref|ZP_00074463.1|
Chla	Chlamydomonas		Green algae	gi|32492888|gb|AAP85534.1|
Sol_IA3	Solanum	IA3	Eudicots	gi|27728149|gb|AAN15319.1|
Ory	Oryza		Monocots	gi|29788240|dbj|BAC75533.1|
Mai1	Maize		Monocots	gi|29126649|gb|AAO17049.2|
Ara2	Arabidopsis		Eudicots	gi|22328517|ref|NP_192641.2|
Ara3	Arabidopsis		Eudicots	gi|25299809|pir||B85091
Syn2	Synechocystis		Cyanobacteria	gi|16329290|ref|NP_440018.1|
Met	Methanosarcina		Euryarchaeotes	gi|23050170|ref|ZP_00077128.1|
Suf	Sulfolobus		Crenarchaeotes	gi|15898878|ref|NP_343483.1|
Nos1	Nostoc		Cyanobacteria	gi|23126243|ref|ZP_00108145.1|
Trid3	Trichodesmium		Cyanobacteria	gi|23039592|ref|ZP_00071150.1|
Art	Arthrobacter		High GC Gram+	gi|7648481|dbj|BAA94842.1|
Mag	Magnetococcus		Proteobacteria	gi|22999082|ref|ZP_00043086.1|
Dei	Deinococcus		Eubacteria	gi|15805296|ref|NP_293987.1|
Chlp	Chlamydophila		Chlamydiaceae	gi|29840170|ref|NP_829276.1|
Chl	Chlamydia		Chlamydiaceae	gi|15604761|ref|NP_219545.1|
Pro1	Prochlorococcus		Cyanobacteria	gi|33240852|ref|NP_875794.1|
Eco	E.coli		Enterobacteria	gi|16131305|ref|NP_417889.1|
Sol_IA2	Solanum	IA2	Eudicots	gi|27728147|gb|AAN15318.1|
Pec	Pectobacterium		Enterobacteria	gi|22074054|gb|AAL02393.1|
Mai2	Maize		Monocots	gi|29126647|gb|AAO17048.3|
Theg	Thermotoga		Thermotogales	gi|15644588|ref|NP_229641.1|
The	Thermus		Eubacteria	gi|3089609|gb|AAC15073.1|
Strp	Streptococcus		Eubacteria	gi|22974081|ref|ZP_00020458.1|
Nos2	Nostoc		Cyanobacteria	gi|23130645|ref|ZP_00112458.1|
Ana1	Anabaena		Cyanobacteria	From CyanoBase
Ana2	Anabaena		Cyanobacteria	From CyanoBase
Pro2	Prochlorococcus		Cyanobacteria	From CyanoBase
Pro3	Prochlorococcus		Cyanobacteria	From CyanoBase
Syc	Synechococcus		Cyanobacteria	From CyanoBase

Table 4. Amino Acids Change of Debranching Enzyme and Plant IA1 and IA3 Clade from its Nearest Ancestor of Isoamylase-type Debraching Enzyme Analysis. Those Amino Acids Shown in Bold Lie within 13 Conserved Regions Identified.

Position	Ancestoral aa	Changed aa
8	L	P
15,16	TE	FK
19	**R**	**S**
30	D	V
33	G	A
43	**E**	**S**
47	**G**	**S**
49	**E**	**T**
51	**L**	**V**
58	D	L
61, 62, 63	PAL	DDV
65	I	D
67	L	M
70, 71, 72	R G E	I P F
74	T	E
124	A	E
134	**V**	**I**
142	G	N
155	A	N
157, 158	EK	QQ
169	D	S
180	I	V
183	S	R
185	I	V
187	D	G
191	A	Q
194	N	P
204	F	V
206, 207, 208	DRD	PCW
210	A	D
213	M	G
219	I	S
226	G	E
258	P	L
271	**L**	**M**
274	K	R
278	K	A
295	P	H
297	G	H
330, 331	QD	FE

Table 4. (continued)

Position	Ancestoral aa	Changed aa
340	F	P
343	D	G
347	T	V
354	**P**	**T**
365	**S**	**A**
370	D	I
375	E	C
383, 384	**DM**	**TL**
410	**L**	**K**
423	**S**	**K**
433, 434	KG	NA
468	**M**	**Y**
485, 486, 487, 488, 489, 490	TNGEVL	HSAWDP
492, 493	GS	QY
528	I	L
546	**T**	**C**
569	G	S
584	**D**	**T**
596	P	F
598	S	G
600	**L**	**F**
605	**L**	**Q**
615	S	T
652	**N**	**A**
694	Q	A
697	**L**	**S**
710	**L**	**F**
732	**F**	**I**
737	**G**	**D**
742	**A**	**T**
747	N	S
757, 758	AD	LE
775	**K**	**N**
790	R	N
796, 797	E T	D Y
800	P	S
844	D	G
855	L	P
857	E	N
865	S	K
866	R	F
868, 869	LA	AF

Table 4. (continued)

Position	Ancestoral aa	Changed aa
871	T	L
873, 874, 875	NED	GKH
897	L	V
913	L	P
916	D	E
918	S	K
923	L	F
927	A	G
938	E	L
945	GE	NS
972	V	I

With regard to cyanobacterial sequences, two *Synechocystis*, three *Trichodesmium*, two *Nostoc*, one *Synechococcus* and three *Prochlorococcus* GDBE sequences were the only IA-like GDBE sequences found and are included in this phylogenetic analysis. With regard to plant sequences, annotations and studies of IA isoforms are only available in *Solanum*, so other plant IAs can only be categorized based on their apparent phylogenetic relationships to the *Solanum* isoforms. Hussain's research showed that the three isoforms fall into structurally distinct isoform classes, so plant IA genes grouped with the potato IA1 isoform all will be considered IA1 type, plant isoamylase genes grouped with the potato IA3 all will be considered IA3 type and plant isoamylase genes grouped with the potato IA2 all will be considered IA2 type. Amino acids changed for each taxon from their nearest ancestors within the cyanobacterial glycogen debranching enzyme, IA1 and IA3 clade according to the character change analysis are listed in Table 4. Those changes are the supportive evidence that cyanobacterial type 1 GDBE, IA1 and IA3 are grouped into one clade (Clade I) and share a close evolutionary relationship.

As was seen in Hussain's research, the three isoforms of plant IA fall into three distinct clades, which suggests distinct evolutionary separation of the isoforms (Hussain et al., 2003). Cyanobacterial sequences fall into four large clades. IA isoforms IA1 and IA3 group together, forming two separate clades within the larger Clade I. One *Synechocystis* and one *Trichodesmium* GDBE isoform (type I GDBEs) fall in the same clade with plant IA1, IA3 and *Chlamydomonas* IA, and Chlamydia and Chlamydophila GDBEs form a sister clade with this Clade I. Plant IA3 isoforms and cyanobacterial type I GDBE isoforms diverge at about

the same time from the line leading to plant IA1 isoforms, suggesting the three groups may have arisen from a common ancestor through gene duplication. The earlier divergence of *Chlamydomonas* isoform IA1 relative to plant IA1 suggests an earlier emergence of the corresponding gene in green algae relative to higher plants. The IA2 isoform, in a clade (Clade II) quite different from Clade I, which includes the IA1 and IA3 isoforms, diverges much earlier than the other two, so seems more distantly related two the other two isoforms. Thus, the two type I cyanobacterial GDBE isoforms appear more closely related to the plant IA1 and IA3 isoforms than are the plant IA2 isoforms.

Besides the IA1/IA3-like GDBE in Clade I discussed above, there are at least three other isoforms of cyanobacteria IA-like GDBE. One isoforms exists in *Synechococcus*, *Trichodesmium*, and *Prochlorococcus* (three duplicates pro1, pro2 and pro3), and *Methanosarcina*. Another isoform exists in *Nostoc*, *Trichodesmium*, scattering amongst the various bacterial GDBE isoforms, such as suf (*Sulfolobus*), mag (*Magnetococcus*), art (*Arthrobacter*), dei (*deinococcus*), eco (*E. coli*) and pec (*Pectobacterium*). The third isoform (Clade III) exists in *Anabaena* (two duplicates, ana1 and ana2) and *Nostoc*. As in the GS case, this may results from either multiple duplication events in ancient bacteria, or multiple duplications in cyanobacteria and lateral gene transfer from cyanobacteria to bacteria. The fact that one particular kind of bacteria or cyanobacteria may not contain all kinds of isoforms could be explained by gene loss during evolution.

The entry of cyanobacteria into eukaryotic cells through endosymbiosis gave rise to the ancestor of plastids. This phylogenetic analysis suggests that plant IA isoforms IA1 and IA3 may be derived from duplication of an ancestral type I (syn1 and trid1) cyanobacterial GDBE gene. IA isoform IA2 appears fairly divergent from the IA1 and IA3 isoforms, as well as from the syn1 and trid1 (type I) cyanobacterial GDBE isoform, so this lineage may have arisen from the ancestral GDBE gene of the host eukaryotic cell or from another IA-like GDBE isoform from the endosymbiont. In this regard, the IA2 clade (Clade II) appears to form a weak sister clade with Clade III, containing multiple cyanobacterial GDBEs. The diversity amongst the cyanobacterial and bacterial IA-like GDBEs, certainly may have provided a rich genetic background from which needed multiple plant IA genes could have arisen. Similar to the evolution of GS genes to SS genes, genes of cyanobacterial origin

would have moved by gene trafficking to the nucleus and become nuclear-encoded genes, even though the functional products are transported back into the plastids to take part in starch biosynthesis or metabolism.

Ancient Ancestral Relationship among Chlamydiaceae, Cyanobacteria, and Chloroplasts

The bacteria in the Chlamydiaceae family, including basceria in the genera *Chlamydia* and *Chlamdophila*, are mainly animal and human pathogens that require intracellular infection of host cells to replicate and survive. As a result of the sequencing of these important pathogens, genes similar to plant gene sequences were found in high proportion (Stephens et al., 1998; Kalman et al., 1999; Read et al., 2000; Shirai et al., 2000). Endosymbiotic bacteria are known to have given rise to mitochondria and plastids, and automated analysis of protein similarity with bacterial proteins demonstrated that *Richettsia*, an α-proteobacterium, proteins have a high similarity to mitochondrial proteins, and *Synechocystis* and *Chlamydiaceae* proteins have high similarity to chloroplast proteins (Brinkman et al., 2002).

The similarity between Chlamydiaceae genes or proteins and those of plants was at first proposed to result from horizontal gene transfer (Stephens et al. 1998; Wolf et al. 1999; Lange et al. 2000; Royo et al. 2000), but this was not supported by analysis of the G+C ratio variance in genes from *Chlamydiaceae* genomes (Brinkman et al. 2002). Observations that the majority of plant-like proteins in *Chlamydiaceae* species correspond to plant genes derived from the plastid (Brinkman et al. 2000; Wolf et al. 1999), and evidence from both plastid ribosomal RNA sequence and intron-splicing analyses suggesting a relationship between *Chlamydiaceae* and plastid lineages (Everett et al. 1999; Nelson et al. 2000) all link the *Chlamydiaceae* with the cyanobacterial/plastid lineage. The details of this evolutionary relationship are unclear, however.

Why starch?

The complete conversion from glycogen synthesis in bacteria and cyanobacteria to starch synthesis in all photosynthetic eukaryotes suggests a strong selective advantage for starch synthesis in these organisms. In considering the question of the origin of starch biosynthesis,

it is important to ask what the selective advantages of starch over glycogen might have been that drove the changes necessary to convert the glycogen synthesizing machinery into starch synthesizing machinery during the evolution of photosynthetic algae and higher plants and what changes were necessary for this to occur.

Glycogen is sometimes thought of as having a simpler structure than starch, but its structure is very important nonetheless. With much more abundant reducing ends than starch, glycogen may represent a glucose storage form with faster mobilization rate than starch. In fact, mathematical models argue that glycogen structure is highly organized and has evolved to satisfy a requirement for rapid glucose mobilization (Meléndez-Hervia et al. 1993, 1995; Meléndez. et al. 1997). Glycogen is a glucose storage polymer in bacteria, cyanobacteria, yeast, animals and humans, all of which may need rapid mobilization of glucose for quick movement and energy mobilization. Plants, on the other hand, normally do not need to make quick movements and, as photosynthetic eukaryotes, have an abundant energy supply through photosynthesis, so they may not have as much demand for rapid glucose mobilization.

In photosynthetic eukaryotes, starch is synthesized to fix the energy from sunlight and degraded to release the energy fixed, as well as to provide the carbon building blocks needed for metabolism and growth. Starch has higher degree of polymerization and parallel arrays of double helices in a complex ordered array formed by interaction between neighboring chains, while glycogen has shorter chains and more branches in a more extended and less packaged structure. For this reason, starch is water insoluble, while glycogen is water soluble. Biochemical reactions take place in water, so, even though the starch molecule is accessible to biochemical action at the surface of the insoluble granule, starch will not be a preferred substrate for rapid glucose mobilization and energy production. Instead, it is a much better storage molecule than an energy supplier, and, because of structural and solubility differences, starch can pack much more glucose and energy in the same volume than can glycogen. This is important in organisms that accumulate carbohydrate in the light for use during the dark and for long-term accumulation on a developmental time scale. The insoluble nature of starch may also be important with regard to its relative inactivity as an osmotic molecule, thus minimizing the need for water of hydrolysis in plants for which water

is one of the most limiting requirements. Thus the ability to store very large amounts of carbohydrate, along with the osmotic inactivity of starch, make starch a much better carbohydrate storage form for plants and other photosynthetic eukaryotes.

CONCLUSIONS

Glycogen Synthesis in Bacteria and Cyanobacteria Gave Rise to Starch Synthesis in Higher Plants

It is well established that plastids arose from a single ancient endosymbiosis between a eukaryotic cell and an ancestral cyanobacterium. Not only the plastidic genes, but also some of the nuclear-encoded genes, may come from the ancient cyanobacterium via gene duplication and trafficking. Glycogen and starch are both glucose storage molecules, and they have many similar, but also different, properties with regard to synthesis, degradation, structure and function. Glycogen is synthesized in bacteria, cyanobacteria, fungi, yeast, animals and humans, while starch only exists in photosynthetic eukaryotes. Cyanobacteria, such as *Synechocystis*, *Nostoc*, *Anabaena*, and *Crocosphaera* have two isoforms of GS, one of which shows very high similarity to one group of starch synthases (SSIII and SSIV). Similarly, one isoform of cyanobacterial IA-like GDBE (type I from *Synechocystis* and *Trichodesmium*) is closely related to two of the three plant IA isoforms. Because cyanobacteria represent the ancestors of ancient plastids, and starch is a photosynthetic product synthesized in plastids, these intertwined relationships suggest that glycogen synthesis in ancestral endosymbiotic cyanobacteria played a direct role in the origin of starch synthesis in ancestral photosynthetic eukaryotes. More specifically, we propose that the combination of glycogen synthesizing and metabolizing enzymes from the endosymbiotic cyanobacterium and the host eukaryotic cell provided the diversity and flexibility needed to begin making a more starch-like glucan. Under selective pressure for the advantages of starch as a glucose storage molecule for a photosynthetic eukaryote, the complement of glycogen synthesizing enzymes present in the early photosynthetic eukaryotes may have begun the transition to starch synthesis. For example, an ancestral cyanobacterial GSI gene may have been duplicated to give rise to starch synthases III, and IV, and an ancestral type I cyanobacterial GDBE gene duplicated to give rise to IA isoforms IA1 and IA3. Both of these gene duplication events may have occurred either before or after transfer of the genes from the ancestral plastid to the nucleus. At the same time, genes involved in glycogen synthesis from the host cell may have contributed, either directly or through duplication, additional components, such as the SSI, SSII and gbSS isoforms, as well as IA isoform IA2.

The lack of any starch-synthesizing organisms yet identified without all five forms of SS or all three forms of IA, suggests that, once the ancestral photosynthetic eukaryotes began to make a more starch-like storage carbohydrate, the conversion was complete and rapid on an evolutionary time scale. It also suggests that all five isoforms of SS and all three isoforms of IA may be required for effective synthesis of starch granules, although it does not rule out formation of intermediate forms of starch-like (e.g., containing some crystalline nature) carbohydrates made by ancestral eukaryotes during the transition.

There are still many questions that need to be answered. Such as whether all those genes not clearly of cyanobacterial origin could have come from the host cell alone. Additional whole genome comparisons may provide at least a partial answer to the questions raised by this proposal.

APPENDIX I CONSEVED SEQUENCE MOTIFS FOR GLYCOGEN SYNTHASES AND STARCH SYNTHASES

```
chl_GS      MKITHTAIEFAPVIKAGGLGDALYGLAKALAVN-HTTEVVIPLY
chlp_GS     MKIIQTAVEFAPFIKAGGLGDAVYGLSKALSES-HDVEVLLPFF
sol_SSIII   MHIVHIAVEMAPIAKVGGLGDVVTSLSRAVQDLNHNVDIILPKY
vig_SSIII   MHIVHIAVEMAPIAKVGGLGDVVTSLSRAVQDLNHNVDIILPKY
ara_S2      LHIVHIAVEMAPIAKVGGLGDVVTSLSRAVQELNHNVDIVFPKY
aeg_SSIII   MRIIHIAVEMAPVAKVGGLGDVVTSLSRAVQDLGHTVEVILPKY
tri_SSIII   MRIIHIAVEMAPVAKVGGLGDVVTSLSRAIQDLGHTVEVILPKY
mai_SSIII   MHIVHIAVEMAPIAKVGGLGDVVTSLSRAVQDLGHNVEVILPKY
chla_S2     LNVVHIASEMAPVAKVGGLADVTASLAKAHQASGLLTEIILPKY
vig_SSV     LYVIHIAAEMAPVAKVGGLGDVISGLSKALQKKGHLVEIILPKY
ara_S1      LYVVHIAAEMAPVAKVGGLGDVVAGLGKALQRKGHLVEIILPKY
tri_SSIV    LHIAHIAAEMAPVAKVGGLADVISGLGKALQKKGHLVEIILPKY
nos_GSI     MYIVQIASECAPVIKAGGLGDVVYGLSRELEIRGNCVELILPKY
ana_GSI     MYIVQIASECAPVIKAGGLGDVVYGLSRELEIRGNCVELILPKY
syn_GSI     MYIVQIASECAPVIKAGGLGDVIYGLSRELELRGHCVELILPMY
cro_GSI     MYIVQVASECAPVIKSGGLGDVVFGLSRELENQGHAVDVILPKY
mag_GSI     MKILMVASECAPVAKVGGLGDVVHGLSRELHQRGHDVAVILPKY
chla_S1     MHIIHITAEMAPIAKVGGLGDVVTGLAKAALARGHFVTVMLPFY
mag_GS      MRVLFVTSEIYPLIKTGGLGDVAAALPAALIEQGVDIRVLVPGY
eco_GS      MQVLHVCSEMFPLLKTGGLADVIGALPAAQIADGVDARVLLPAF
dei_GS      MRVLHLASEVFPFSRSGGLGDVLGALPAVQARLGEDAEVTVLSP
nos_GSIIa   MRILFVAAEAAPIAKVGGMGDVVGALPKVLRKMGHDVRIFLPYY
ana_GSII    MRILFVAAEAAPIAKVGGMGDVVGALPKVLRKMGHDVRIFLPYY
nos_GSIIb   MRILFVAAEAAPIAKVGGMGDVVGALPKFLREMGHDVRIFLPYY
ths_GSIIa   MRILFVAAEAAPLAKVGGMGDVVGSLPKVLRRMGHDVRIFMPYY
ths_GSIIb   MRILFVAAEAAPLAKVGGMGDVVGSLPKVLRRMGHDVRIFMPYY
syn_GSII    MKILFVAAEVSPLAKVGGMGDVVGSLPKVLHQLGHDVRVFMPYY
cro_GSII    MRILFVGAEVAPLAKAGGMGDVLGALPKVLHQLGYDVRIIMPYY
syc_GSII    MRILFVAAECAPFAKVGGMGDVVGSLPKVLKASAHDVGIFMRYY
pro_GSII    MRVLFAAAECAPMVKVGGMGDVVASLPSALAKLGHDVRLIIPGY
ara_S3      MNVILVAAECAPFSKTGGLGDVAGALPKSLARRGHRVMVVVPRY
man_bSII    MNVILVAAECAPWSKTGGLGDVAGSLPKALARRGHRVMVVAPRY
sol_bSII    MNIILVASECAPWSKTGGLGDVAGALPKALARRGHRVMVVAPRY
pea_bSII    MNIILVSAECAPWSKTGGLGDVAGSLPKALARRGHRVMIVAPHY
aeg_SSII    MNVVVVAAECSPWCKTGGLGDVAGALPKALAKRGHRVMVVVPRY
tri_SSII    MNVVVVAAECSPWCKTGGLGDVAGALPKALAKRGHRVMVVVPRY
hor_SSII    MNVVVVAAECSPWCKTGGLGDVAGALPKALAKRGHRVMVVVPRY
mai_SSII    MNVIVVAAECSPWCKTGGLGDVVGALPKALARRGHRVMVVVPRY
chla_S1     MNVVMVGAECAPWSKTGGLGDVMAALPKALVRRGHRVMVVVPRY
tri_SSI     RSIVFVTGEAAPYAKSGGLGDVCGSLPIALAARGHRVMVVMPRY
aeg_SSI     RSIVFVTGEAAPYAKSGGLGDVCGSLPIALAARGHRVMVVMPRY
sor_SS      QSIVFVTGEASPYAKSGGLGDVCGSLPVALAARGHRVMVVMPRY
ory_SS      RSVVFVTGEASPYAKSGGLGDVCGSLPIALALRGHRVMVVMPRY
ara_S4      NNLVFVTSEAAPYSKTGGLGDVCGSLPIALAGRGHRVMVISPRY
ara_S5      MSVIFIGAEVGPWSKTGGLGDVLGGLPPALAARGHRVMTICPRY
sor_bS      MNLIFVGTEVGPWSKTGGLGDVLGGLPPALAARGHRVMTISPRY
pea_bS      MSLVFVGAEVGPWSKTGGLGDVLGGLPPVLAGNGHRVMTVSPRY
hor_bS      MPIIFVATEVHPWCKTGGLGDVVGGLPPALAAMGHRVMTIAPRY
ory_bS      MNVVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVISPRY
mai_bS      MNVVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVVSPRY
tri_bS      MNLVFVGAEMAPWSKTGGLGDVLGGLPPAMAANGHRVMVISPRY
chla_bS     LDIVMVAAEVAPWSKTGGLGDVTGGLPIELVKRGHRVMTIAPRY
```

```
chl_GS      DDAFRFCAFSAAAAAY
chlp_GS     DDTLRFSAFSAAAAAY
sol_SSIII   NDGERFGFFCHAALEF
vig_SSIII   NDAERFGFFCHAALEF
ara_S2      DDAGRFGFFCHAALEF
aeg_SSIII   NDDRRFGFFCHSALEF
tri_SSIII   NDDRRFGFFCHSALEF
mai_SSIII   DDDRRFGFFCRSALEF
chla_S2     DDGARFSFFSRAALEF
vig_SSV     DDFRRFSYFSRAALEF
ara_S1      DDFRRFSYFSRAALEL
tri_SSIV    DDFKRFSYFSRAALEL
nos_GSI     DDDMRFAFFSKAALEF
ana_GSI     DDDMRFAFFSKAALEF
syn_GSI     DDHMRFAFFSKAAMEF
cro_GSI     DDNLRFAFFSRAVLEY
mag_GSI     DDAHRFGFFSRAALEF
chla_S1     NEMEAYLYFCRACLEY
mag_GS      DNAIRFAALSRVAALI
eco_GS      DNVLRFALLGWVGAEM
dei_GS      DDVERFCAFGRAALPA
nos_GSIIa   DEDWRFTLFSNGAAEF
ana_GSII    DEDWRFTLFSNGAAEF
nos_GSIIb   DEDWRFTFFSNGAAEF
ths_GSIIa   DEDWRFTFFANGAAEF
ths_GSIIb   DEDWRFTFFANGAAEF
syn_GSII    DEAWRFTFFSNGAAEF
cro_GSII    DEYWRFTFFSNAAAEF
syc_GSII    DEDWRFTFFANGAAEF
pro_GSII    DEDWRFTFFASATAEF
ara_S3      DILKRMVLFCKAAVEV
man_bSII    DILKRMVLFCKAAVEV
sol_bSII    DILKRMVLFCKAAIEV
pea_bSII    DILRRMVLFCKAAVEV
aeg_SSII    EIMKRMILFCKAAVEV
tri_SSII    EIMKRMILFCKAAVEV
hor_SSII    EIMKRMILFCKAAVEV
mai_SSII    EIMKRMILFCKVAVEV
chla_S1     EILFRCALLCKAALEA
tri_SSI     DNQFRYTLLCYAACEA
aeg_SSI     DNQFRYTLLCYAACEA
sor_SS      DNQFRYTLLCYAACEA
ory_SS      DNQFRYTLLCYAACEA
ara_S4      DNQFRFTLLCHAACEA
ara_S5      DNQLRFSLLCQAALEA
sor_bS      DNELRFSLLCQAALEA
pea_bS      DNQLRFSLLCQAALEA
hor_bS      DNQLRFCLLCLAALEA
ory_bS      DNQMRFSLLCQ---EA
mai_bS      DNQLRFSLLCQAALEA
tri_bS      DNQLRFSLLCQAALEA
chla_bS     DNHKRFALFCKAAIEA
```

```
chl_GS      HDWHVGLV
chlp_GS     HDWHMGLL
sol_SSIII   HDWSSAPV
vig_SSIII   HDWSSAPV
ara_S2      HDWSSAPV
aeg_SSIII   HDWSSAPV
tri_SSIII   HDWSSAPV
mai_SSIII   HDWSSAPV
chla_S2     HDWQAAAV
vig_SSV     HDWQTAFI
ara_S1      HDWQTAFV
tri_SSIV    HDWQTAFV
nos_GSI     HDWQTGLI
ana_GSI     HDWQTGLI
syn_GSI     HDWQTGLV
cro_GSI     HDWQTGLL
mag_GSI     HDWQTALI
chla_S1     HDWHAAAA
mag_GS      NDWQAGLA
eco_GS      HDWHAGLA
dei_GS      HDWQAGLV
nos_GSIIa   HDWHTGMI
ana_GSII    HDWHTGMI
nos_GSIIb   HDWHTGMI
ths_GSIIa   HDWHTGMI
ths_GSIIb   HDWHTGMI
syn_GSII    HDWHTGMI
cro_GSII    HDWHTGMI
syc_GSII    HDWHTGMI
pro_GSII    HDWHTGMI
ara_S3      NDWHTALL
man_bSII    NDWHTALL
sol_bSII    NDWHTALL
pea_bSII    NDWHTALL
aeg_SSII    NDWHTALL
tri_SSII    NDWHTALL
hor_SSII    NDWHTALL
mai_SSII    NDWHTALL
chla_S1     NDWHTALL
tri_SSI     NDWHASLV
aeg_SSI     NDWHASLV
sor_SS      NDWHASLV
ory_SS      NDWHASLV
ara_S4      NDWHAGLV
ara_S5      NDWHTALL
sor_bS      NDWHTALI
pea_bS      NDWHSALI
hor_bS      NDWHTAVL
ory_bS      NDWHTGPL
mai_bS      NDWHTGPL
tri_bS      NDWHTGLL
chla_bS     NDWHSALV
```

```
chl_GS      LTLHNFGYRG
chlp_GS     FTIHNFSYRG
sol_SSIII   FTIHNLEFG-
vig_SSIII   FTIHNLEFG-
ara_S2      FTIHNLEFG-
aeg_SSIII   FTIHNLEFG-
tri_SSIII   FTIHNLEFG-
mai_SSIII   FTIHNLEFG-
chla_S2     LTIHNMAFQG
vig_SSV     FTCHNFEYQG
ara_S1      FTCHNFEYQG
tri_SSIV    FTCHNFEYQG
nos_GSI     YTIHNFKHQG
ana_GSI     YTIHNFKHQG
syn_GSI     YTIHNFKHQG
cro_GSI     YTIHNFKHQG
mag_GSI     HTIHNFKHQG
chla_S1     LTIHNLDNTG
mag_GS      ITVHNIQFQG
eco_GS      FTVHNLAYQG
dei_GS      FSVHNLQYQG
nos_GSIIa   FTIHNLAYQG
ana_GSII    FTIHNLAYQG
nos_GSIIb   FTIHNLAYQG
ths_GSIIa   FTIHNLAYQG
ths_GSIIb   FTIHNLAYQG
syn_GSII    FTIHNLAYQG
cro_GSII    FTIHNLAYQG
syc_GSII    FTIHNLAYQG
pro_GSII    FTIHNLKYQG
ara_S3      LVIHNIAHQG
man_bSII    LVIHNIAHQG
sol_bSII    LVIHNIAHQG
pea_bSII    LVIHNIAHQG
aeg_SSII    MVIHNIAHQG
tri_SSII    MVIHNIAHQG
hor_SSII    MVIHNIAHQG
mai_SSII    LVIHNIAHQG
chla_S1     FVIHNMAHQG
tri_SSI     LVIHNLAHQG
aeg_SSI     LVIHNLAHQG
sor_SS      LVIHNLAHQG
ory_SS      LVIHNLAHQG
ara_S4      LIIHNLAHQG
ara_S5      FCIHNIAYQG
sor_bS      FCIHNIAYQG
pea_bS      FCIHNIAYQG
hor_bS      FCIHNIAYQG
ory_bS      FCIHNISYQG
mai_bS      FCIHNISYQG
tri_bS      FCIHNISYQG
chla_bS     LAIHNIAFQG
```

```
chl_GS      DFVTTVSPTYAKEI
chlp_GS     DYITTVSPTYAQEI
sol_SSIII   DKATTVSPTYSQEV
vig_SSIII   DKATTVSPTYSREI
ara_S2      DKATTVSPTYAKEV
aeg_SSIII   DKATTVSPTYSRDV
tri_SSIII   DKATTVSPTYSRDV
mai_SSIII   DKATTVSNTYSKEV
chla_S2     DRITTVSPTYANEV
vig_SSV     NIVTTVSPTYAQEV
ara_S1      NIVTTVSPTYAQEV
tri_SSIV    NIVTTVSPTYALEV
nos_GSI     NAVTTVSPNHALEA
ana_GSI     NAVTTVSPNHALEA
syn_GSI     NYVNTVSPHHAWEA
cro_GSI     NAVNTVSPHHAWEA
mag_GSI     NFVTTVSPKHAFEA
chla_S1     NAVTTVSPTYANEV
mag_GS      NWLTTVSPTYANEI
eco_GS      DHITAVSPTYAREI
dei_GS      GHVTTVSPTYAQEI
nos_GSIIa   DRVNTVSPTYAEQI
ana_GSII    DRVNTVSPTYAEQI
nos_GSIIb   NKVNTVSPTYAEQI
ths_GSIIa   DRVNTVSPTYAEQI
ths_GSIIb   DRVNTVSPTYAEQI
syn_GSII    NRVTTVSPTYAQQI
cro_GSII    GRITTVSPTYAEQI
syc_GSII    DRVNTVSPTYAQQI
pro_GSII    DRVNAVSPTYSREI
ara_S3      DRVLTVSHGYSWEV
man_bSII    DRVVTVSHGYAWEL
sol_bSII    DRVVTVSHGYSWEL
pea_bSII    DRIVTVSHGYAWEL
aeg_SSII    DQVVVVSPGYLWEL
tri_SSII    DQVVVVSPGYLWEL
hor_SSII    DQVVVVSPGYLWEL
mai_SSII    DRVVTVSRGYLWEL
chla_S1     HRLVAVSKCYAWEC
tri_SSI     DRIVTVSQGYSWEV
aeg_SSI     DRIVTVSQGYSWEV
sor_SS      DRIVTVSKGYSWEV
ory_SS      DRIVTVSQGYSWEV
ara_S4      DRIITVSQGYAWEI
ara_S5      HRVLTVSPYYAQEL
sor_bS      HRVVTVSPYYAQEL
pea_bS      DQVFTVSPHYAKEL
hor_bS      DVVLTVSPHYVKEL
ory_bS      DRVLTVSPYYAEEL
mai_bS      DRVLTVSPYYAEEL
tri_bS      DKVLTVSPYYAEEL
chla_bS     DKLVTVSPNYATEI
```

```
chl_GS      GILNGIDYTILDPETD
chlp_GS     GILNGIDEQIWNPETD
sol_SSIII   GIVNGIDPDIWDPLND
vig_SSIII   GIINGIDPDIWDPFND
ara_S2      GIINGIDPDIWDPYND
aeg_SSIII   GILNGIDPDIWDPYTD
tri_SSIII   GILNGIDPDIWDPYTD
mai_SSIII   GILNGIDPDIWDPYND
chla_S2     GVLNGLDYETWNPATD
vig_SSV     GILNGIDTDIWNPATD
ara_S1      GILNGIDTDSWNPATD
tri_SSIV    GILNGIDTDTWNPCTD
nos_GSI     GVLNGIDYDFWNPEID
ana_GSI     GVLNGIDYDFWNPEID
syn_GSI     GILNGLDYEVWNPEID
cro_GSI     GILNGLDYEVWNPEVD
mag_GSI     GILNGLDYDTWNPMTD
chla_S1     GILNGIDCEEWNPATD
mag_GS      GILNGIDYHAWNPETD
eco_GS      GVLNGVDEKIWSPETD
dei_GS      GILNGLDQDRWNPRTD
nos_GSIIa   GIVNGIDTEVYDPAND
ana_GSII    GIVNGIDTEVYDPAND
nos_GSIIb   GIINGIDTEVYNPEDD
ths_GSIIa   GILNGIDTELFDPSSD
ths_GSIIb   GILNGIDTELFDPSSD
syn_GSII    GILNGIDTEIYNPAED
cro_GSII    GILNGIDMELYNPAED
syc_GSII    GILNGIDVDSYNPATD
pro_GSII    GILNGIDLDEWDPATD
ara_S3      GIVNGIDTQEWNPEFD
man_bSII    GIVNGIDAKEWNPQFD
sol_bSII    GIVNGIDTKEWNPELD
pea_bSII    GIVNGVDTKDWNPQFD
aeg_SSII    GIVNGIDNMEWNPEVD
tri_SSII    GIVNGIDNMEWNPEVD
hor_SSII    GIVNGIDNMEWNPEVD
mai_SSII    GIVNGIDHQEWNPKVD
chla_S1     GIVNGIDYKEWNPICD
tri_SSI     GIVNGIDINDWNPTTD
aeg_SSI     GIVNGIDINDWNPTTD
sor_SS      GIVNGIDINDWNPATD
ory_SS      GIVNGIDINDWNPSTD
ara_S4      GITNGINVDEWNPSTD
ara_S5      GIINGMDVQEWNPSTD
sor_bS      GIVNGMDTQEWNPATD
pea_bS      GIVNGMDNREWSPQTD
hor_bS      GIVNGMDVIDWNPATD
ory_bS      GIVNGMDVSEWDPSKD
mai_bS      GIVNGMDVSEWDPSRD
tri_bS      GIVNGMDVSEWDPAKD
chla_bS     GIVNGMDIEEWNPKTD
```

```
chl_GS      KAENKKALYETLGL
chlp_GS     KEENKNALYEKLGL
sol_SSIII   KTAAKEALQRKLGL
vig_SSIII   KRASKEALQQKLGL
ara_S2      KRAAKEELQNRLGL
aeg_SSIII   KRAAKRALQQKFGL
tri_SSIII   KRAAKRALQQKFGL
mai_SSIII   KRAAKRALQQKFGL
chla_S2     KRACKAALLRELGL
vig_SSV     KSENKEALRRNLGL
ara_S1      KEENKHALRKQLGL
tri_SSIV    KAANKAALREQLNL
nos_GSI     KLYNKKALRERLLL
ana_GSI     KLYNKKALRERLLL
syn_GSI     KAKNKQALRERLLL
cro_GSI     KDLNKKGLRERLWL
mag_GSI     KSDNKQALRERLML
chla_S1     KALCKEFLQKGLGL
mag_GS      KLLNKRALQQEMGL
eco_GS      KAENKRQLQIAMGL
dei_GS      PAGKAGAVKALRQE
nos_GSIIa   RKANKIALQEEVGL
ana_GSII    RKANKIALQEEVGL
nos_GSIIb   RKANKIALQEEVGL
ths_GSIIa   RRANKIALQEELGL
ths_GSIIb   RRANKIALQEELGL
syn_GSII    RVKNKIAIQEETGL
cro_GSII    RKANKIGLQQETGL
syc_GSII    RLNNRLALQKEMGL
pro_GSII    RKKNKEALQRQMGL
ara_S3      KPQCKAALQKELGL
man_bSII    KPQCKTALQNELRF
sol_bSII    KPQCKAALQKELGL
pea_bSII    KRQCKAALQRELGL
aeg_SSII    KRQCKEALQRELGL
tri_SSII    KRQCKEALQRELGL
hor_SSII    KRQCKEALQRELGL
mai_SSII    KRQCKAALQRELGL
chla_S1     KAKCKAALQKELGL
tri_SSI     KAKCKAELQKELGL
aeg_SSI     KAKCKAELQKELGL
sor_SS      KAKCKSALQKELGL
ory_SS      KAKCKAELQKELGL
ara_S4      KIKCKMALQKELGL
ara_S5      KPLIKEALQAAVGL
sor_bS      KPLLKEALQAAVGL
pea_bS      KPLLKGTLQAEIGL
hor_bS      RALNKEILQAEVGL
ory_bS      KALNKEALQAEAGL
mai_bS      KALNKEALQAEVGL
tri_bS      KALNKEALQAEVGL
chla_bS     KAAAKEALQAELGL
```

```
chl_GS      PCMCIISRIAEQKGPEFMKQAILHAL
chlp_GS     PLMCIISRIVEQKGPEFMKAAILHAM
sol_SSIII   PLVGIITRLTHQKGIHLIKHAIWRTL
vig_SSIII   PLVGVITRLTHQKGIHLIKHAIWRTL
ara_S2      PVVGIITRLTHQKGIHLIKHAIWRTL
aeg_SSIII   PIVGIITRLTAQKGIHLIKHAIHRTL
tri_SSIII   PIVGIITRLTAQKGIHLIKHAIHRTL
mai_SSIII   PVVGIVTRLTAQKGIHLIKHAIHRTL
chla_S2     PLIAIVSRLTQQKGLHLMKAGLKAAV
vig_SSV     PLVGCITRLVPQKGVHLIRHAIYLTL
ara_S1      PLVGCITRLVPQKGVHLIRHAIYRTL
tri_SSIV    PLVGCITRLVAQKGVHLIRHAIYKTA
nos_GSI     PIIAYIGRLDNQKGVHLVHHAIYHAL
ana_GSI     PIIAYIGRLDNQKGVHLVHHAIYHAL
syn_GSI     PMLCFIGRLDGQKGVHLVHHSIYYAL
cro_GSI     PIIAYIGRLDDQKGVGLVLHALEYAL
mag_GSI     PMVAYVGRLDTQKGVHLIRHALFRTL
chla_S1     PLVAVVSRLVPQKGIHLIKAALFRTV
mag_GS      PLFGLVSRLTEQKGIDLILEAIPAVL
eco_GS      PLFAVVSRLTSQKGLDLVLEALPGLL
dei_GS      PILATVSRLADQKGMDLLITALPELV
nos_GSIIa   FLIGMVTRLVEQKGLDLVIQMLDRFM
ana_GSII    FLIGMVTRLVEQKGLDLVIQMLDRFM
nos_GSIIb   FLIGIVTRLVEQKGIDLILQILDRFL
ths_GSIIa   FLVGMVSRLVEQKGLDLLIQILDRFL
ths_GSIIb   FLVGMVSRLVEQKGLDLLIQILDRFL
syn_GSII    MVVGIVARLVEQKGIDLVIQILDRFM
cro_GSII    FLMGMVTRLVEQKGLDLVLQMLDRFL
syc_GSII    FLIGFVARLVEQKGIDLLLQILDRFL
pro_GSII    YLLGMVGRLVDQKGVDLLLQVSRRLL
ara_S3      PLIGFIGRLDHQKGVDLIAEAVPWMM
man_bSII    PVIGFIGRLDYQKGVDLIAEAIPWMV
sol_bSII    PLIGFIGRLDPQKGVDLIAEAVPWMM
pea_bSII    PIISFIGRLDHQKGVDLIAEAIPWMM
aeg_SSII    PLLGFIGRLDGQKGVEIIADAMPWIV
tri_SSII    PLLGFIGRLDGQKGVEIIADAMPWIV
hor_SSII    PLLGFIGRLDGQKGVEIIADAMPWIV
mai_SSII    PLLGFIGRLDGQKGVDIIGDAMPWIA
chla_S1     PMLGFIGRLDYQKGVDLIRDNYDYIM
tri_SSI     PLIGFIGRLDYQKGIDLIKMAIPELM
aeg_SSI     PLIGFIGRLDYQKGIDLIKMAIPELM
sor_SS      PLIGFIGRLDYQKGIDLIQLIIPHLM
ory_SS      PLIGFIGRLDYQKGIDLIKLAIPDLM
ara_S4      PMIGFIGRLDYQKGIDLIQTAGPDLM
ara_S5      PVIGFIGRLEEQKGSDILVEAISKFM
sor_bS      PLIGFIGRLEEQKGSDILVAAIHKFI
pea_bS      PLIGFIGRLEEQKGSDILVEAIAKFA
hor_bS      PVIVFIGRLEEQKGSDILIAASPEFV
ory_bS      PLIAFIGRLEEQKGPDVMAAAIPELM
mai_bS      PLVAFIGRLEEQKGPDVMAAAIPQLM
tri_bS      PLVAFIGRLEEQKGPDVMIAAIPEIL
chla_bS     PLFAFIGRLEEQKGVDIILAALPKIL
```

```
chl_GS      TLIIIGTC
chlp_GS     ALVIVGTC
sol_SSIII   QVVLLGSA
vig_SSIII   QVVLLGSA
ara_S2      QVVLLGSA
aeg_SSIII   QVVLLGSA
tri_SSIII   HVVLLGSA
mai_SSIII   QVVLLGSA
chla_S2     QVVVLGTA
vig_SSV     QFVLLGSS
ara_S1      QFVLLGSS
tri_SSIV    QFVLLGSS
nos_GSI     QFVLLGSA
ana_GSI     QFVLLGSA
syn_GSI     QFVLLGSA
cro_GSI     QFVLLGSA
mag_GSI     QFVLLGAA
chla_S1     QFVLLGSG
mag_GS      QLVVLGSG
eco_GS      QLALLGAG
dei_GS      -VVVLGGG
nos_GSIIa   QFVLLGTG
ana_GSII    QFVLLGTG
nos_GSIIb   QFVLLGTG
ths_GSIIa   QFVLLGTG
ths_GSIIb   QFVLLGTG
syn_GSII    QLIILGTG
cro_GSII    QLVVLGSG
syc_GSII    QFVVLGTG
pro_GSII    QIVVLGTG
ara_S3      QLVMLGTG
man_bSII    QLVMLGTG
sol_bSII    QLVMLGTG
pea_bSII    QLVMLGTG
aeg_SSII    QLVMLGTG
tri_SSII    QLVMLGTG
hor_SSII    QLVMLGTG
mai_SSII    QLVMLGTG
chla_S1     QLVMLGSG
tri_SSI     QFVMLGSG
aeg_SSI     QFVMLGSG
sor_SS      QFVMLGSG
ory_SS      QFVMLGSG
ara_S4      QFVMLGSG
ara_S5      QMVILGTG
sor_bS      QIVVLGTG
pea_bS      QIVVLGTG
hor_bS      QIIVLGTG
ory_bS      QIVLLGTG
mai_bS      QIVLLGTG
tri_bS      QIVLLGTG
chla_bS     QIAILGTG
```

```
chl_GS      LTYSDVLARQIFAAADMICIPSMFEPCGLTQMIGMRYGTVPVVRATGGLADTV
chlp_GS     LDYNDPLARLVYGAADMICIPSHFEPCGLTQLIGMRYGTVPLVRSTGGLADTV
sol_SSIII   LTYDEPLSHLIYAGADFILVPSIFEPCGLTQLTAMRYGSIPVVRKTGGLYDTV
vig_SSIII   LAYDEPLSHMIYAGADFILVPSIFEPCGLTQLTAMRYGSIPIVRKTGGLYDTV
ara_S2      LTYDEPLSHLIYAGADFILVPSIFEPCGLTQLIAMRYGAVPVVRKTGGLFDTV
aeg_SSIII   LTYDEPLSHLIYAGSDFIIVPSIFEPCGLTQLVAMRYGSIPIVRKTGGLYDTV
tri_SSIII   LTYDEPLSHLIYAGSDFIIVPSIFEPCGLTQLVAMRYGSIPIVRKTGGLHDTV
mai_SSIII   LTYDEPLSHLIYAGSDFILVPSIFEPCGLTQLVAMRYGTIPIVRKTGGLFDTV
chla_S2     LRHDEAMSHRIFAASDMLLVPSMFEPCGLTQLIALRYGTVPVVRETGGLADTV
vig_SSV     LKYDESLSHAIYAASDMFIIPSIFEPCGLTQMISMRYGAIPIARKTGGLNDSV
ara_S1      LKYDEALSHTIYAASDLFIIPSIFEPCGLTQMIAMRYGSIPIARKTGGLNDSV
tri_SSIV    LKYDDALSHCIYAASDMFIVPSIFEPCGLTQMIAMRYGSVPIVRKTGGLNDSV
nos_GSI     LGFNEELSHLIYAGADMIVVPSNYEPCGLTQMIGLKYGTVPIVRGVGGLVNTV
ana_GSI     LGFNEELSHLIYAGADMIVVPSNYEPCGLTQMIGLKYGTVPIVRGVGGLVNTV
syn_GSI     LGFDEELAHLIYGAADIIVVPSNYEPCGLTQMIGLRYGAVPVVRGVGGLVNTV
cro_GSI     LSFNEELSHLIYAGADMMVVPSNYEPCGLTQMISLKYGTVPIVRGVGGLVNTV
mag_GSI     IGFDEALAHQIYAGVDLFVVPSLFEPCGLTQMIALRYGTVPVVRQVGGLADTV
chla_S1     IMYSERLAHMIYAAADVVVVPSMFEPCGLTQMIALRYGAVPLVRRTGGLADTV
mag_GS      IGYDEIHSHRIQAGVDILLVPSRFEPCGLTQLYAMRYGTLPLVRRTGGLADSV
eco_GS      IGYHEAFSHRIMGGADVILVPSRFEPCGLTQLYGLKYGTLPLVRRTGGLADTV
dei_GS      SGMNEPLAHRIYAGAHAFAMPSRFEPCGLSQLIAMRYGTLPIVRETGGLADTV
nos_GSIIa   LLYNDALSRRIYAGTDAFLMPSRFEPCGISQMMALRYGSIPIVRRTGGLVDTV
ana_GSII    LLYNDALSRRIYAGTDAFLMPSRFEPCGISQMMALRYGSIPIVRRTGGLVDTV
nos_GSIIb   LLYNDALSRRIYAGTDAFLMPSRFEPCGISQMMSLRYGSVPIVRRTGGLVDTV
ths_GSIIa   LLYSDVLSRRIYGGADAFIMPSRFEPCGISQMIALRYGCVPIVRRTGGLVDTV
ths_GSIIb   LLYSDVLSRRIYGGADAFIMPSRFEPCGISQMIALRYGCVPIVRRTGGLVDTV
syn_GSII    LLHNDALSRRVYAGADVFLMPSRFEPCGLSQLMAMRYGCIPIVRRTGGLVDTV
cro_GSII    ILYNDVLSRRIFAGADVFLMPSRFEPCGISQLFAMRYGCVPIVRSTGGLKDTV
syc_GSII    LMYDEGLSRRIYAGSDAFLVPSRFEPCGITQMLALRYGSVPIVRRTGGLVDTV
pro_GSII    LTYDDYLSRLIYAGSDAFLMPSRFEPCGISQLLAMRYGSIPIVRNVGGLVDTV
ara_S3      VGFSVKTAHRITAGADILLMPSRFEPCGLNQLYAMNYGTIPVVHAVGGLRDTV
man_bSII    VGFSVKTAHRITAGADILLMPSRFEPCGLNQLYAMMYGTIPVVHAVGGLRDTV
sol_bSII    VGFSVKTSHRITAGADILLMPSRFEPCALNQLYAMKYGTIPVVHAVGGLRDTV
pea_bSII    VGFSVKMAHRITAGSDILLMPSRFEPCGLNQLYAMSYGTVPVVHGVGGLRDTV
aeg_SSII    VGFSVRLAHRITAGADALLMPSRFEPCGLNQLYAMAYGTVPVVHAVGGLRDTV
tri_SSII    VGFSVRLAHRITAGADALLMPSRFEPCGLNQLYAMAYGTVPVVHAVGGLRDTV
hor_SSII    VGFSVRLAHRITAGADALLMPSRFEPCGLNQLYAMAYGTIPVVHAVGGLRDTV
mai_SSII    VGFSVPMAHRITAGADVLVMPSRFEPCGLNQLYAMAYGTVPVVHAVGGLRDTV
chla_S1     VGFSNKMAHRITAAADILLMPSRFEPCGLNQLYAMAYGTVPIVHSVGGLRDTV
tri_SSI     VGFSVPVSHRITAGCDILLMPSRFEPCGLNQLYAMQYGTVPVVHGTGGLRDTV
aeg_SSI     VGFSVPVSHRITAGCDILLMPSRFEPCGLNQLYAMQYGTVPVVHGTGGLRDTV
sor_SS      VGFSVPVSHRITAGCDILLMPSRFEPCGLNQLYAMQYGTVPVVHATGGLRDTV
ory_SS      VGFSVPVSHRITAGCDILLMPSRFEPCGLNQLYAMQYGTVPVVHGTGGLRDTV
ara_S4      VGFNVPISHRITAGCDILLMPSRFEPCGLNQLYAMRYGTIPVVHGTGGLRDTV
ara_S5      AKFNVPLAHMITAGADFIIVPSRFEPCGLIQLHAMRYGTVPIVASTGGLVDTV
sor_bS      AKFNVPLAHMITAGADFMLVPSRFEPCGLIQLHAMRYGTVPICASTGGLVDTV
pea_bS      TKFNSPLAHKIIAGADFIVIPSRFEPCGLVQLHAMPYGTVPIVSSTGGLVDTV
hor_bS      AKFNVPLAHMMFAGADFIIIPSRFEPCGLIQLQGMSYGVVPICSSTGGLVDTV
ory_bS      VKFNAPLAHLIMAGADVLAVPSRFEPCGLIQLQGMRYGTPCACASTGGLVDTV
mai_bS      VKFNAALAHHIMAGADVLAVTSRFEPCGLIQLQGMRYGTPCACASTGGLVDTI
tri_bS      VRFNAPLAHQMMAGADVLAVTSRFEPCGLIQLQGMRYGTPCACASTGGLVDTI
chla_bS     VKFSAPLAHMLTAGADFMLVPSRFEPCGLIQLHAMHYGTVPVVASTGGLVDTV
```

```
chl_GS      IVRSCLEFSSDLETAANKYLEI
chlp_GS     LVEEGMLRSSGLTTMAIHYLGV
sol_SSIII   LCKQVMEQDWSWNRPALDYLEL
vig_SSIII   LCKTVMEQDWSWNRPALDYLEL
ara_S2      LCKTVMEQDWSWNRPALEYLEL
aeg_SSIII   LCKRVMEQDWSWNRPALDYIEL
tri_SSIII   LCKRVMEQDWSWNRPALDYIEL
mai_SSIII   LCKRVMEQDWSWNRPALDYIEL
chla_S2     LMPRIMRQDFGWTQSASTYQRI
vig_SSV     LVQKDMNIDFSWDSSAAQYEEL
ara_S1      LVEKVMSIDFSWGSSATQYEEL
tri_SSIV    LVQKDMTIDFSWDTSASQYEDI
nos_GSI     LAIQGMKYDYSWNNPGTEYLNI
ana_GSI     LAIQGMKYDYSWNNPGTEYLNI
syn_GSI     LALQGMAYDYSWNKPGLQYVEA
cro_GSI     LAIQGMRCDYSWRKPVEQYIEV
mag_GSI     LMQAGMRYDYSWRLPVEHYLNI
chla_S1     LQQDNMRLDVSWGKSAKSYVDV
mag_GS      AQQRAMSMDFGWRKSARAYMHL
eco_GS      VQRQAMAMDFSWQVAAKSYREL
dei_GS      RATRAMSLDFSWDGPARQYLEL
nos_GSIIa   LQKRGMSQDFSWYKSAKEYDKL
ana_GSII    LQKRGMSQDFSWYKSAKEYDKL
nos_GSIIb   LQKRGMSEDFSWYKSAKEYVKL
ths_GSIIa   LQQRGMREDFSWTKSALAYNAL
ths_GSIIb   LQQRGMREDFSWTKSALAYNAL
syn_GSII    LQQRAMRADFSWYRSAGEYIKV
cro_GSII    LQQRCMTQDFSWYGSAAEYIKI
syc_GSII    LQQRGMAVDLSWKQSAIAYEQL
pro_GSII    LQKRAMTQMYSWERSAMEYETM
ara_S3      LQRRGMTQDLSWDNAAEKYEEV
man_bSII    LQRRGMTPNLSWDHAAEKYEET
sol_bSII    IQTRCMTQDLSWDNAAQNYEEV
pea_bSII    IQERGMSQDLSWDNAAQQYEEV
aeg_SSII    LQERGMSQDFSWEHAAKLYEDV
tri_SSII    LQERGMSQDFSWEHAAKLYEDV
hor_SSII    LQERGMSQDFSWEHAAKLYEDV
mai_SSII    LQARGMSQDLSWDHAAELYEDV
chla_S1     IQRRGMEQDLTWDNAASIYEEV
tri_SSI     LMKRGMTKDHTWDHAAEQYEQI
aeg_SSI     LMKRGMTKDHTWDHAAEQYEQI
sor_SS      AHEEGHVKRLHVGPCR------
ory_SS      SNEARHVKRLYMGPCRLTV---
ara_S4      LMRRGMTRNYSWENAAVQYEQV
ara_S5      MVKNCMDQDFSWKGPARLWEKV
sor_bS      MIKNCMSEELSWKEPAKKWETL
pea_bS      IILNCMAQNFSWKKPAKLWEKA
hor_bS      MVQNCMAQDLSWKGPAKKWEEA
ory_bS      MVRNCMNQDLSWKGPAKNWENV
mai_bS      MVRNCMIQDLSWKGPAKNWENV
tri_bS      MVKNCMIQDLSWKGPAKNWEDV
chla_bS     MVANCISQDLSWSKPAQKWEGL
```

APPENDIX II CONSEVED SEQUENCE MOTIFS FOR GLYCOGEN DEBRANCHING ENZYMES AND ISOAMYLASES

theg	DIDVTNYVRIVLSESLKEE
the	GETVKYVCRFASAFEFGAT
syn1	GVNFSIYSSHSTACTLVLF
trid1	GVNFSIFSSYATSCTLVLF
sol_IA1	GVNFAVFSRNATAATLCLI
ara1	GVNFSVYSTNSVSATICLI
tri	GVNFAVYSGGATAAALCLF
aeg	GVNFAVYSGGATAAALCLF
hor	GVNFAVYSGGATAAALCLF
ory	GVNFAVYSAGASAASLCLF
chla	AINFSVFSSSAESVSLVLF
sol_IA3	GINFAIFSQHASAVTLCII
mai1	GLNFAIFSQHASSVTLCLN
ara2	GINFALFSQNATSVTLCLS
chlp	RYRFALFSSQATQVVLALS
chl	RYRFSLFSSQAQQVTLVLL
trid2	GVNFSIYSKNATAIELLLF
syn2	GVNFCLFSKHAERVTLLLF
met	GVNFSIYSEHADYVELLLF
pro1	GVNFSVAAPNASYVELLLF
pro2	GVNFSLIATNAEYIEILLF
pro3	GVNFSVAAPTAKRVELLLF
syc	GVNFSVAAPAADRIELLIF
nos1	GTNFALFSENATGVELCLF
trid3	GTNFAIFSENATSVELCLF
art	GTNFALFSEHAEKVELCLF
dei	GTNFALYSENATGVELCLF
mag	GVNFAIFSEHAERVELCLF
eco	GVNFTLFSAHAERVELCVF
pec	GVNFALFSSGASRVELCIF
sol	GVNFSLFSENAEKVELLLY
ara3	SWNFSFFSRSSTNVVLCLY
sol_IA2	SVNFALFSRSARSVVLCLY
mai2	AANFAVYSKIAKGMVLCLF
nos2	LTDFKLFAPRNKGAALIGS
ana1	PIEFTLFAPYNKGAALIAS
strp	NIKITDKAGKTVAIDELTL
ana2	ECEFTVWSPTLNSVAVQIL

```
theg      DGELGAVYSPE
the       DGPLGAEYLKE
syn1      GNVYCMVVFDL
trid1     GNVYCMIVFDL
sol_IA1   TGDVWHVFLKG
ara1      TGHVWHVFLRG
tri       TGNVWHVFIEG
aeg       TGNVWHVFIEG
hor       TGDVWHVFLEG
ory       TGNVWHVFIEG
chla      TGDVWHIMLPD
sol_IA3   TGDIWHICIKE
mai1      TGDIWHVSVEG
ara2      TGDTWHICVED
chlp      TGAVWHIEVEG
chl       TGAIWHIEIAG
trid2     THHYWHIFISG
syn2      TFYYWHVFVKG
met       TFHFWHVYVRG
pro1      SGDYWHIEVEG
pro2      TGSYWHAEINN
pro3      SGDYWHVEVEG
syc       SGDYWHVELEG
nos1      NNFVWHAYLPG
trid3     SNFIWHGYLPG
art       DGYVWHCYLPQ
dei       TAFVWHGYLPG
mag       TNQVWHGYLPD
eco       SGDIWHGYLPD
pec       TGDIWHGYLPD
sol       TGDIWHVFVPG
ara3      TGDVWHASVDN
sol_IA2   SGDIWHAALDC
mai2      TGDVWHVSMES
nos2      SEDGYFRTKIK
ana1      GDDGYFRTTVE
strp      DGELGATLAKD
ana2      GEGYWQIKVND
```

```
theg      SPVSKWV
the       APTATAV
syn1      YGYRMEG
trid1     YGYRMDG
sol_IA1   YGYKFDG
ara1      YGYRFDG
tri       YGYRFDG
aeg       YGYRFDG
hor       YGYRFDG
ory       YGYRFDG
chla      YGYRVEG
sol_IA3   YGYRIDG
mai1      YGYRING
ara2      YGYRVDG
chlp      YAFRVDG
chl       YAYKLRG
trid2     YAYRVYG
syn2      YAYRVDG
met       YAYRIGG
pro1      YAYRVYG
pro2      YAFRVK-
pro3      YGYRIFG
syc       YGYRVFG
nos1      YGFRVHG
trid3     YGFRVHG
art       YGYRVHG
dei       YGYRVHG
mag       YGYRVYG
eco       YGYRVHG
pec       YGYRVDG
sol       YAYRVYG
ara3      YGYRCKE
sol_IA2   YGYRCKA
mai2      YGFRSGL
nos2      YKFRIQT
ana1      YKFRVQT
strp      WSPSADS
ana2      YRYQLND
```

```
theg      SPVSKWV
the       APTATAV
syn1      YGYRMEG
trid1     YGYRMDG
sol_IA1   YGYKFDG
ara1      YGYRFDG
tri       YGYRFDG
aeg       YGYRFDG
hor       YGYRFDG
ory       YGYRFDG
chla      YGYRVEG
sol_IA3   YGYRIDG
mai1      YGYRING
ara2      YGYRVDG
chlp      YAFRVDG
chl       YAYKLRG
trid2     YAYRVYG
syn2      YAYRVDG
met       YAYRIGG
pro1      YAYRVYG
pro2      YAFRVK-
pro3      YGYRIFG
syc       YGYRVFG
nos1      YGFRVHG
trid3     YGFRVHG
art       YGYRVHG
dei       YGYRVHG
mag       YGYRVYG
eco       YGYRVHG
pec       YGYRVDG
sol       YAYRVYG
ara3      YGYRCKE
sol_IA2   YGYRCKA
mai2      YGFRSGL
nos2      YKFRIQT
ana1      YKFRVQT
strp      WSPSADS
ana2      YRYQLND
```

```
theg      IIYEIHIADIT
the       IIYELSIRDFT
syn1      VIYEMHVRGFT
trid1     IIYEMHVRSFT
sol_IA1   VIYEMHVRGFT
ara1      VIYEMHVRGFT
tri       VIYEMHLRGFT
aeg       VIYEMHLRGFT
hor       VIYEMHLRGFT
ory       VIYEMHLRGFT
chla      VIYEAHVRGFT
sol_IA3   VIYEMNVRAFT
mai1      VIYEMNVRAFT
ara2      VIYEMNVRAFT
chlp      IIYEMHVRSFT
chl       FIYEMHVRSFT
trid2     VIYEMHVGGFT
syn2      IIYELHVGGFT
met       IIYELHVGGFT
pro1      IIYELHVGGFT
pro2      IIYELHIKAFT
pro3      VIYELHVGGFS
syc       VIYELHVGGFT
nos1      IIYETHVRGFT
trid3     VIYEINLKGFT
art       VIYEAHVKGLT
dei       VIYEAHVKGLT
mag       IIYEMHVRGFT
eco       IIYEAHVKGLT
pec       VLYEAHVRGLT
sol       VIYEVHVKGFT
ara3      LVYRLNVKGFT
sol_IA2   IIYRLNVTQFT
mai2      VVYRANVALFT
nos2      VIYEMHVADFT
ana1      VIYELHVGDFS
strp      IIYEAHVRDFT
ana2      IFYELHVGTFT
```

```
theg      LSHLVELGVTHVHILPFFDF
the       LSYVKELGVTHVQLMPFMDF
syn1      IPYLQELGVNTIELMPIFEF
trid1     IPYLKELGVNAVELMPIYEF
sol_IA1   LDHLKELGVNCIELMPCHEF
ara1      LDHLKELGINCIELMPCHEF
tri       LDYLKELGVNCIELMPCHEF
aeg       LDYLKELGVNCIELMPCHEF
hor       LDYLKELGVNCIELMPCHEF
ory       LDYLKELGVNCVELMPCHEF
chla      LDYLKSLGVNAIELLPVFEF
sol_IA3   IPHLLELGVNAVELLPVFEF
mai1      IPHLLELGVNAVELLPVFEF
ara2      IPHLQDLGINAVELLPVFEF
chlp      IDYLKKLGINAIELLPIFEF
chl       IDHLKQLGVHAVELLPIFEF
trid2     IPYLKELGITAVELLPVQQF
syn2      IPYLKELGITAVELLPVHYF
met       IPYLKELGITAVELMPVFDF
pro1      IPYLKDLGITTIELLPVFAF
pro2      IPYLKKLGITSIELLPVFCF
pro3      LPYLRKLGITAIELLPIFAF
syc       LPYLKELGITAIELLPVFCF
nos1      IQYLQQLGITSVELMPVHHF
trid3     IAYLQSLGITAVELMPIHHF
art       ISHLQKLGVTAIELMPVHQF
dei       LDYLRDLGITAIEFLPVHQH
mag       ISYLQDLGISSVELLPVHAS
eco       INYLKQLGITALELLPVAQF
pec       LDYLTQLGVTALELMPVQQH
sol       ISYLKDLGITTVELMPVFHF
ara3      VSHLKTLGTNAVLLEPIFSF
sol_IA2   WHHFKDLGVNAMLLEPIFPF
mai2      VEHFRHLGVNAVLLEPVFPF
nos2      LDYLCELGINAIELMPVNEY
ana1      LDYLCELGINAIELLPVKEY
strp      LDYLKDLGVTHVQLLPVLSY
ana2      LPELRELGINAIELMPIAQF
```

```
theg      WGYDPYLFMVPEGRYST
the       WGYNPLHLYAPEGSYAT
syn1      WGYSTVNFFAPKAGYAA
trid1     WGYSTVGFFAPKAGYAA
sol_IA1   WGYSTVNFFSPMGRYSS
ara1      WGYSTIGFFSPMIRYAS
tri       WGYSTINFFSPMTRYTS
aeg       WGYSTINFFSPMTRYTS
hor       WGYSTINFFSPMTRYTS
ory       WGCSTINFFSPMIRYSS
chla      WGYSTVNYFSPMGRFSA
sol_IA3   WGYSTINFFAPMSRYAS
mai1      WGYSTINFFAPMSRYAS
ara2      WGYSTVNFFAPMSRYAS
chlp      WGYASVNFFSPCRRYAY
chl       WGYSSVNFFCPSRRYTY
trid2     WGYSQIAFFAPHHSYSS
syn2      WGYSTIGFFAPHQGYSA
met       WGYDPICFFAPHSGYCV
pro1      WGYSPVNWFTPHQSFIS
pro2      WGYSPINWFTPHFQYLS
pro3      WGYSPLNWFTPHPKYVH
syc       WGYSPLSWFTPHHGYGC
nos1      WGYDSINYFTPYSGYSA
trid3     WGYDSINYFAPYSGYSS
art       WGYNTIGFFAPQNTYSS
dei       WGYSTLNFFAPDVRYSA
mag       WGYNTLAFFVANPRFAS
eco       WGYNPVAMFALHPAYAC
pec       WGYNTLLPFAVDNSLAA
sol       WGYDPINFFSPECRYSS
ara3      GPYFPFHFFSPMDIYGP
sol_IA2   GPYFPWHFFSPGNMYGP
mai2      GPYFPYHFFSPMSLYSS
nos2      WGYKVRHFFATESSYGS
ana1      WGYNPRYFFATESSYGS
strp      WGYDPQHYFALSGMYSA
ana2      WGYDGVYPFAVQNSYGS
```

```
theg      EVKEMVKALHKHGIGVIMDMVFPHTYGIGELSA
the       ELKQAIHTLHENGLRVVMDAVYNHVYDR-EQSP
syn1      ELKNLVKELHKVGISVILDVVFNHTAEGNERGP
trid1     ELKTLVKDLHKNGIEVILDVVFNHTAEGNENGP
sol_IA1   EFKYLVKEAHKRGIEVIMDVVFNHTAEGNENGP
ara1      EFKILVKEAHKRGIEVIMDVVLNHTAEGNEKGP
tri       EFKTFVREAHKRGIEVILDVVFNHTAEGNENGP
aeg       EFKTFVREAHKRGIEVILDVVFNHTAEGNENGP
hor       EFKTFVRESHKRGIEVILDVVFNHTAEGNENGP
ory       EFKTFVREAHKRGIEVIMDVVFNHTAEGNEKGP
chla      EFKQLVKECHRRGIEVILDVVFNHTAEGNERGP
sol_IA3   EFKEMVKALHGAGIEVILDVVYNHTNEADDENP
mai1      ELKQMVKAFHNSGIEVILDVVYNHTNEADDVNP
ara2      EFKEMVKALHSAGIEVILDVVYNHTNEADDKYP
chlp      EFKTLVKALHKAGIELILDVVFNHTGLDNTTCP
chl       EFKTLVKALHRAGIEVILDVVFNHTGFEGTSCP
trid2     EFRDMVKALHKEGIEVILDVVFNHTAEGNENGP
syn2      EFRDMVKALHKAGIEVILDVVFNHTAEGNEKGP
met       EFRDMVRALHKAGIEVVLDVVFNHTAEGDNLGP
pro1      QFRKLVATCHDNGIEIILDVVYNHTTEGNENGP
pro2      EFRKFVEECHKANIEVILDVVYNHTSEGDYQGP
pro3      QVRELVAACHDEGIEVILDVVYNHTTEGSIDGP
syc       QVRQLVAACHDANIEVLLDVVYNHTTEGTRLGP
nos1      EFKQMVKDLHSAGIEVILDVVYNHTGEGNHLGP
trid3     EFKQMVKALHNGGIEVILDVVYNHTGEGNHLGP
art       DFKAMVRSLHRAGIEVILDVVYNHTAEGNHLGP
dei       EFKNMVRALHDAGIEVILDVVYNHTAEGNHMGP
mag       EFKTMVQVFHEAGIEVLLDVVYNHTAEGNHEGP
eco       EFRDAIKALHKAGIEVILDIVLNHSAELDLDGP
pec       EFRDTVRALHQAGIEVILDVVFNHSAELDVDGP
sol       SFKKMVNELHNAGIEVIIDVVYNHTAEGNHLGP
ara3      SMKVMVKKLHSEGIEVLLEVVFTHTADS-----
sol_IA2   SMKDMVKKLHANGIEVFLEVVFTHTAED-----
mai2      SMKDMVKTMHRNGIEVLLEVVFTHTAEGGAECQ
nos2      DLKRLIDECHGRGIRVFMDGIYNHTDEECPLIL
ana1      DLKKLVDECHQRGIRIIMDGIYNHSEASSPLTQ
strp      ELKNLVNEIHKRGMGVIFDVVYNHTARTYLFED
ana2      DLKNFVNACHENGIAVVLDVVYNHFGPEGNYMG
```

```
theg      YLNESGCGNVIASERPMMRKFIVDTVTYWVKEYHIDGFRFDQMGLIDK
the       PANGTGVGNDIASERRMARRWIVDSVVFWAKEYGIDGFRFDLMGVHDI
syn1      YFNFSGTGNTLNCNNPIVRGMVLDCLRYWTAEFHIDGFRFDLASILGR
trid1     YFNFSGCGNTINCNNPIVRNVVLDCLRYWASEYHIDGFRFDLASILGR
sol_IA1   FYNYSGCGNTFNCNNPIVRQFIVDCLRYWVTEMHVDGFRFDLASILTR
ara1      FYNYSGCGNTFNCNHPVVRQFILDCLRYWVTEMHVDGFRFDLGSIMSR
tri       FYNYSGCGNTFNCNHPVVRQFIVDCLRYWVTEMHVDGFRFDLASIMTR
aeg       FYNYSGCGNTFNCNHPVVRQFIVDCLRYWVMEMHVDGFRFDLASIMTR
hor       FYNYSGCGNTFNCNHPVVRQFIVDCLRYWVMEMHIDGFRFDLASIMTR
ory       FYNYSGCGNTFNCNHPVVREFIVDCLRYWVTEMHVDGFRFDLASIMTR
chla      YYNYSGCGNTLNCNQPVVRQFILDCLKHWVTEYHVDGFRFDLASILTR
sol_IA3   LLNFAGCGNTFNCNHPTVMELILESLRHWVTEYHVDGFRFDLASVLCR
mai1      LLNFSGCGNTLNCNHPVVKELVLDSLRHWVKEYHIDGFRFDLASVLCR
ara2      LLNFSGCGNTLNCNHPVVMELILDSLRHWVTEYHVDGFRFDLASVLCR
chlp      FANYSGCGNTVNTNYTPTTQWILDSLRYWVQEMHVDGFRFDLASVFSR
chl       LMNFSGCGNTVNTNTPTTLKWILDALRYWVQEMHVDGFRFDLASVFSR
trid2     YSNYTGCGNTFNTNNPFVHRLIVDCLCYWVREMHVDGFRFDLASVMSR
syn2      YSNYSGCGNSVKANHPVVGGLILDSLRYWVSEMHVDGFRFDLASVLVR
met       YSNYSGCGNTVSCNHPISQKLIVDCLKYWAKEMHVDGFRFDEGSILSL
pro1      FLDVTGCGNTIAANQPIVRQLILESIKCWSQELGVDGFRFDLGVALSR
pro2      YQDVSGCGNTIAANRGLVRKLILESLKCWVNELGVDGFRFDLGIALSR
pro3      YLDVSGCGNSIAANRPIVRQLILESMRCWAIELGVDGFRFDLGIALSR
syc       YLDVSGCGNSIAANQPISTQLILESMRCWALELGVDGFRFDLGIALSR
nos1      YMDFTGCGNSLNVRHAQVLKLIMDSLRYWVTEMHIDGFRFDLASALAR
trid3     YMDFTGCGNSLNVRHPQILKLIMDSLRYWVTEMHVDGFRFDLASALAR
art       YMDYTGTGNTLNVRQPHSLQLLMDSLRYWVTEMHVDGFRFDLAAALAR
dei       YFDYTGTGNSLNVRHPQTLQLIMDSLRYWVTEMHVDGFRFDLASTLAR
mag       YRDFTGCGNSFNLRHPKVLQLVMDSLRYWAGEMHVDGFRFDLTTTLAR
eco       YHNWTGCGNTLNLSHPAVVDYASACLRYWVETCHVDGFRFDLAAVMGR
pec       YHNWAGCGNVLRLEHPAVLHWVIECLTFWHEVCHVDGFRFDLATILGR
sol       YLDFTGTGNTLNLSHPRVIQMVLDSLRYWVTEMHVDGFRFDLAAALAR
ara3      -ANDLDSKSYLNCNYPVVQQLVLESLRYWVTEFHVDGFCFINASSLLR
sol_IA2   -GQYLNIQNALNCNYPIVQQMILDCLRHWVIEFHIDGFVFVNASSLLR
mai2      -GIAGCKASVLNCNHPVTQKLILDSLRHWVLDFHVDGFCFINAPFLVR
nos2      EFNYDFYDKNLNIQP--AWEYIGDVVRFWIQEYHIDGIRFDAVRQLAN
ana1      EFNYEHYDENLETYP--ARKFIGDTVRYWVGEYHLDGIRYDAARQIAN
strp      TARESFGGGRLGTTHAMSRRILVDSITYLTREFKVDGFRFDMMGDHDA
ana2      KTPWGNAMNFDDAYSQGVRNYFIQNALYWLGEFHIDGLRLDAIQAIYD
```

```
theg      --THVAAFNDEFRDAIRG
the       QLPRFAYFNDRFRDAVKG
syn1      WAEWNGKYRDTVRKFIKG
trid1     WAEWNGKYRDAIRKFIKG
sol_IA1   WSEWNGKYRDMVRQFIKG
ara1      WSEWNGKFRDVVRQFIKG
tri       WSEWNGKYRDIVRQFIKG
aeg       WSEWNGKYRDIVRQFIKG
hor       WSEWNGKYRDIVRQFIKG
ory       WSEWNGKYRDIVRQFIKG
chla      WSEWNGKFRDVVRNFIKG
sol_IA3   WAEWNGKYRDDIRRFIKG
mai1      WAEWNGKYRDDIRRFIKG
ara2      WAEWNGMYRDDVRRFIKG
chlp      WSEWNGQYRDTIKSFLNG
chl       WSEWNGCYRDHVKAFLNG
trid2     FAEWNGHYRDDVRQFVKS
syn2      FAEWNGPFRDDVRRFVKG
met       WAEWNGYYRDEIRRFVRG
pro1      VSTWNGHFRDDLRRFWKG
pro2      TFTWNGHFRDDLRRFWKG
pro3      IGTWNGHYRDDLRAFWKG
syc       IGTWNGHFRDGLRRFWKG
nos1      WSEWNGRYRDTVRDFWRG
trid3     WSEWNGKYRDTVRDFWRG
art       WTEWNGKYRDTVRDFWRG
dei       WAEWNGIYRDDMRSFWKG
mag       WGEWNDRYRDAIRRYWKG
eco       FAEWNDHFRDAARRFWLH
pec       FAEWNDRFRDDMRRFWLH
sol       WAEWNGKYRDSIRRFWRG
ara3      WAELNTRYCRNVRNFLRG
sol_IA2   WAEINMRFCDDIRDFLRG
mai2      WAEMNTRFSMDVRKFLKG
nos2      WLTKQAKNHAAPKQFYNI
ana1      WIAQEAKKTAGAKPFYNV
strp      ATNTVGVFSDDIRNTLKS
ana2      DAQWSDDFHHALHALLTG
```

```
theg      KGFVMGGYG---
the       RGFALGNPG---
syn1      MAQRLQGSPDLY
trid1     MAQRIQGSPDLY
sol_IA1   FAECLCGSPNLY
ara1      FAECLCGSPNLY
tri       FAECLCGSPHLY
aeg       FAECLCGSPHLY
hor       FAECLCGSPQLY
ory       FAECLCGSPHLY
chla      FASAICGSPNIY
sol_IA3   FATRIAGSADLY
mai1      FATRVSGSADLY
ara2      FATRVSGSSDLY
chlp      FASRISGSQDLY
chl       FASRISGSHDIY
trid2     LAARIMGSPDIY
syn2      LASRLLGSPDIY
met       VASRIAGSPDLY
pro1      LKDRLLGSPSIY
pro2      MSDKIKGSPSHY
pro3      MGQRLSGSADLY
syc       LAQRFKGSPDLY
nos1      FAYCFTGSPDLY
trid3     FAYRFTGSSDLY
art       FASRITGSADLY
dei       IGYRITGSSDLY
mag       LASRLTGSSDIF
eco       FAGRFAASSDVF
pec       LARRFAASSEVF
sol       IANRLLGSPDIY
ara3      LATRICGSGDVF
sol_IA2   LATRLCGSGDIF
mai2      LATRLCGSGDLF
nos2      TVVKPDGPLDAC
ana1      SITNLDGPMDGC
strp      AAFITGGAKNLE
ana2      FGKCADLAKAYA
```

```
theg      TKIKRGVVGSINYDGKLIKSFALDPEETINYAACHDNHTLWDKNYLAAKA
the       EQVKLAIAGSLRALG----GLFCHPRQSINYVECHDNHTFWDKMEAANHD
syn1      PSTSINFVTAHDGFTLADLVAYNGKHNYANGENGNDGANDNYSWNCGVEG
trid1     PATSINFITAHDGFTLADLVSYNDKHNEANGENNNDGANDNESWNCGAEG
sol_IA1   PWNSINFVCAHDGFTLADLVTYNNKHNLANGEDNKDGENHNNSWNCGEEG
ara1      PWHSINFICAHDGFTLADLVTYNNKNNLANGEENNDGENHNYSWNCGEEG
tri       PWHSINFVCAHDGFTLADLVTYNNKYNLPNGENNRDGENHNLSWNCGEEG
aeg       PWHSINFVCAHDGFTLGDLVTYNNKYNLPNGENNRDGENHNLSWNCGEEG
hor       PWHSINFVCAHDGFTLADLVTYNNKYNLPNGEDNRDGENHNLSWNCGEEG
ory       PWHSINFVCAHDGFTLADLVTYNKKYNSSNGEDNRDGENHNLSWNCGEEG
chla      PHASINFVAAHDGFTLADMVAYNNKHNEANGENNRDGEQHNNSWNCGEEG
sol_IA3   PYHSVNFVIAHDGFTLYDLVSYNNKHNDANGEGGNDGCNDNFSWNCGIEG
mai1      PYHSVNFVIAHDGFTLCDLVSYNSKHNDANGEGGRDGCNDNYSWNCGIEG
ara2      PYHGVNFVIAHDGFTLRDLVSYNFKHNEANGEGGNDGCNDNHSWNCGFEG
chlp      PCNSINYICSHDGFTLHDTVSYNSKHNEENGEENRDGSNANYSYNFGEEG
chl       PTNSINYICSHDGFTLYDTVAYNDKHNEENGEYNRDGTSANYSYNFGCEG
trid2     PNHSINFITCHDGFTMNDLVSYNEKHNEANCEDNRDGANHNFSWNCGIEG
syn2      INRSINFVTCHDGFTLVDLVSYNEKHNEANGEKNRDGTNDNFSWNCGVEG
met       PINSVNFVTCHDGFTLNDLVSYNHKHNEANGENNRDGIENNLSWNCGVEG
pro1      VEKSINFITSHDGFTLIDLVSFDKKHNLSNGESNRDGENHNNSWNHGVEG
pro2      LPKSINFITSHDGFTLKDLVTFNKKHNFANKEQNRDGDNHNNSWNHGVEG
pro3      LGRSLNFITAHDGFTLNDLVSFNRKHNLANGECNRDGENHNNSWNNGIEG
syc       LGRSVNLITAHDGFTLADLVAYNRKHNLANGEDNRDGENHNNSWNHGVEG
nos1      PSASINFITAHDGFTLNDLVSYNEKHNQDNGEDSRDGESHNRSWNCGVEG
trid3     PHASINFVTAHDGFTLNDLVSYNEKHNEANGEDNNDGEKHNRSWNSGEEG
art       PVASINFVTAHDGFTLRDLVSYNEKHNEANGEDNKDGESHNRSWNCGVEG
dei       PYASINFVTAHDGFTLRDSVTYEQKHNEANGEGNNDGHNHNITWNCGVEG
mag       PWSSINFVTAHDGFTLRDLVSYNHKHNSDNLEDNRDGRDANDSWNCGVEG
eco       PSAAINLVTAHDGFTLRDCVCFNHKHNEANGEENRDGTNNNYSNNHGKEG
pec       PWASVNMLTSHDGFTLRDLVCFNHKHNDANGEQNRDGTNSNFSFNHGTEG
sol       PFASINYVTSHDGFTLEDLVSYNQKHNEANGFNNQDGMNENYSWNCGAEG
ara3      PAFSFNYISRNSGLSLVDIVSFSG-----------PELASELSWNCGEEG
sol_IA2   PAFSFNYIARNSGLTLVDLVSFSS-----------NEVASELSWNCGQEG
mai2      PAFSFNYVSRNSGLTLVDLVSFSS-----------DELASEFSWNYGEEG
nos2      RYFLVPYICGKSFELEQLKQILDPKQQGYAIATNVINYLATHDRERLLRE
ana1      YHTIKAHICGDTFDLENLKDVIDPKRQGFLGATNVVNYLTNHDHDHIMVE
strp      PGDVVQYIAAHDNLTLHDVIAKSINKDPKVAEEEIHKRIRLGNTMILTAQ
ana2      PHRKRFHGISCRDRPLSQFSVCIQNHDQIGNQMQGERLSERISFAGLKLA
```

```
theg      YHKGLIKLRKEHPAF
the       YVQGLIALRRAHGAF
syn1      FVKHCIAFRLAHPVL
trid1     FVKNCVAFRKAHPIL
sol_IA1   FCGLMTKFRHECESL
ara1      FCRILIKFRDECESL
tri       FCCLMTKFRKECEGL
aeg       FCCLMTKFRKECEGL
hor       FCCLMTKFRKECEGL
ory       FCSLMTKFRKQCESL
chla      FVRLLIHFRRATPAL
sol_IA3   FFSKMIKFRLSHNVL
mai1      FFSEMIKFRHNHPIL
ara2      FFSEVIKFRHSHHVL
chlp      FVCNAIRFRKQHKEI
chl       FLCQVIALRKAYTEL
trid2     FVKKIIHFTQELQIT
syn2      FLRGIIALTQSLSLF
met       FWKLMIDFRKRHTTI
pro1      FVKNLISIRKNLSEF
pro2      YLKYLIKIRKKFINF
pro3      FVSRLLMIRHQLSEL
syc       FLQRLLKLRQALPQL
nos1      FARELIYFRHQHPVF
trid3     FTRELIYFRFQHPVF
art       FTAAVNSLRAKHPTF
dei       FTRKLIALRKAHPSL
mag       FTRFLIKLRREHGVF
eco       FTAALIHLRKRIPAL
pec       FTSGLIRLRRSIPAL
sol       FVKKMIQFYRAHPAF
ara3      FISFMTSVRARRSDV
sol_IA2   FISFLSNLRMRRSDL
mai2      FISFLSALRSRRADI
nos2      YYQKLITLRQQTPAL
ana1      YHKGLIGLRKNNHAL
strp      YTQGLIALRRSTDAF
ana2      WYRQLIHLRKTHPAL
```

REFERENCES CITED

Altschul, S. (1990). A basic local alignment search tool. Journal of molecular biology. 215: 403-410.

Archibald, JM., Keeling, PJ. (2002). Recycled plastids: a 'green movement' in eukaryotic evolution. Trends in Genetics. 18(11): 577-584.

Baldauf, SL. et al. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. Science. 290: 972-977.

Ball, SG., Morell, MK. (2003). Annual Review of Plant Biology. 54: 207-33.

Beatty, MK., Rahman, A., Cao, HP., Woodman, W., Lee, M., Myers, AM., and James, MG. (1999). Purification and molecular genetic characterization of ZPU1, a pullulanase-type starch-debranching enzyme from maize. Plant Physiology. 119: 255-266.

Black, RC., Loerch, JD., McArdle, FJ., and Creech, RG. (1996). Genetic interactions affecting maize phytoglycogen and the phytoglycogen-forming branching enzyme. Genetics. 53: 661-668.

Brinkman, FSL., Blanchard, JL., Cherkasov, A., Av-Gay,Y., Brunham, RC., Fernandez, RC., Finlay, BB., Otto, SP., Francis Quellette, BF., Keeling, PJ., Rose, AM., Hancock, REW., Jones, SJM. (2002). Evidence that plant-like genes in *Chlamydia* Species reflect an ancestral relationship between Chlamydiaceae, Cyanobacteria, and chloroplast. Genome Research. 12: 1159-1167.

Burton, R., Jenner., Carrangis, L., Fahy, B., Fincher, G., Hylton, C., Laurie, D., Parker, M., Waite, D., van Wegen, S., Verhoeven, T., and Denyer, K. (2002). Starch granule initiation and growth are altered in barley mutants that lack isoamylase activity. Plant Journal. 31: 97-112.

Cao H., Radosevich, I., Guan, H., Keeling, PL., James, MG., Myers, AM. (1999). Identification of the soluble starch synthase activities of maize endosperm. Plant Physiology. 120: 205-215.

Commuri, PD, Keeling, PL. (2001). Chain-length specificities of maize starch synthase I enzyme: studies of glucan affinity and catalytic properties. Plant Journal. 25:475-486.

Dinges, JR., Colleoni, C., James, MG., Myers, AM. (2003). Mutational analysis of the pullulanase-type debranching enzyme in maize indicates multiple functions in starch metabolism. Plant Cell. 15:666-680.

Douglas, SE., Penny, SL. (1999). The plastid genome from the cryptomonad alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. Journal of Molecular Evolution. 48: 236-244.

Edwards, A., Borthakur A., Boremann, S., Venail, J., Denyer, K., Waite, D., Fulton, D., Smith, A., Martin, C. (1999). Specificity of Starch Synthase Isoforms from Potato. European Journal of Biochemistry. 166: 724-736.

Everett, KD., Kahane, S., Bush, RM., and Friedman, MG. (1999). An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. Journal of Bacteriology. 181: 4734-4740.

Flügge, U-I., Heldt, HW. (1991). Metabolite translocators of the chloroplast envelope. Annual Review of Plant Physiology and Plant Molecular Biology 42:129-144.

Fujita, N., Kubo, A., Francisco, PB., Nakakita, M., Harada, K., Minaka, N., and Nakamura, Y. (1999). Purification, characterization and cDNA structure of isoamylase from developing endosperm of rice. Planta. 208: 283-293.

Fontaine, T., D'Hulst, C., Maddelein, M-L., Routier, F., Marianne-Pepin, T., et al. (1993). Toward and understanding of the biogenesis of the starch granule. Evidence that Chlamydomonas sobuble starch synthase II controls the synthesis of intermediate size glucans of amylopectin. Journal of Biological Chemistry. 268: 16223-30.

Furukawa, K., Tagaya, M., Inouye, M., Preiss, J., Fukui, T. (1990). Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. Journal of Biological Chemistry. 265: 2086-2090

Furukawa, K., Tagaya, M., Tanizawa, K., Fukui, T. (1993). Role of the conserved Lys-x-Gly-Gly sequence of the ADP-glucose binding site in Escherichia coli glycogen synthase. J. Biol. Chem. 269: 23837-23842.

Gao, M., Wanat, J., Stinard, PS., James MG., Myers, AM. (1998). Characterization of dull1, a maize gene coding for a novel starch synthase. Plant Cell. 10: 399-412.

Goff, SA., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). Science. 296: 92-100.

Gray, MW. (1992). The endosymbiont hypothesis revisited. International Review of Cytology. 141: 233-357.

Guan, HP. and Preiss, J. (1993). Differentiation of the properties of the branching isozymes from maize (*Zea mays*). Plant Physiology. **102**: 1269–1273.

Hall, BG. (2001). Phylogenetic trees made easy. 40-76. Inauer Associates, Inc. Publishers. Sunderland, Massachusetts. USA

Honda, D., Yokota, A., Sugiyama, J. (1999). Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine synechococcus strains. Journal of Molecular Evolution. 48: 723-739.

Hussain, H., Mant, A., Seale, R., Zeeman, S., Hinchliffe, E., Edwards, A., Hylton, C., Bornemann, S., Smith, A., Martin, C., Bustos, R. (2003). Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. The Plant Cell,. 15: 133-149.

James, MG., Robertson, DS., Myers, AM. (1995). Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell. 7: 47-429.

James, MG., Denyer, K., Myers, AM. (2003). Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology. 6: 215-222

Jenkins, PJ., Cameron, RE., Donald, AM. (1993). A universal feature in the structure of starch granules from different botanical sources. Starch. 45: 405-409.

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyuman, RW., Olinger, L., Grimwood, J., Davis, RW., and Stephens, RS. (1999). Comparative genomes of *Chlamydia* pneumoniae and C. trachomatis. Nature Genetics. 21: 385-389

Kaneko, T., Tabata, S. (1997). Complete genome structure of the unicellular cyanobacterium *Synehocycsis* sp. PCC6803. Plant Cell Physiology. 38(11): 1171-6.

Keeling, PL., Wood, JR., Tyson, RH., Bridges, IG. (1998). Starch biosynthesis in the developing wheat grain. Evidence against the direct involvement of triose phosphate in the metabolic pathway. Plant Physiology. 87: 311-319.

Kubo, A., Fujita, N., Harada, K., Matsuda, T., Satoh, H., and Nakamura, Y. (1999). The starch-debranching enzymes isoamylase and pullulanase are both involved in amylopectin biosynthesis in rice endosperm. Plant Physiology. 121: 399-409.

Lange, BM. Rujan, T., Martin, W., and Croteau, R., (2000). Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences. 92:13172-13177.

Levi, C., Gibbs, M. (1976). Starch degradation in isolated chloroplasts. Plant Physiology. 57: 933-935.

Madison, WP. (1989). Reconstructing character evolution on polytomous cladograms. Cladistics. 5:365-377.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacteria and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academy of Sciences. 99(19): 12246-12251.

Marshall, J., Sidebottom, C., Debet, M., Martin, C., Smith, AM., Edwards, A. (1996). Identification of the major starch synthase in the soluble fraction of potato tubers. Plant Cell. 8: 1121-35.

Martin, C., Smith, AM. (1995). Starch biosynthesis. Plant Cell. 7: 971-85.

Meléndez,R., Meléndez-Hevia, E., Cascante, M. (1997). How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. Journal of Molecular Evolution. 45: 446-455.

Meléndez-Hevia, E., Waddell, T.G., Shelton, DE. (1993). Optimization of molecular design in the evolution of metabolism: the glycogen molecule. Biochemistry Journal. 295: 477-483.

Mizuno, K., Kobayashi, E., Tachibana, M., Kawasaki, T., Fujimura, T., Funane, K., Kobayashi, M., Baba, T. (2001). Characterization of an isoform of starch branching enzyme, RBE4, in developing seeds. Plant and Cell Physiology. 42(4): 349-357.

Moreira, D. et al. (2000). The origin of red algae and the evolution of chloroplasts. Nature. 405: 69-72

Morrison, WR., Milligan, TP., Azudin, MNJ. (1984). Cereal Science. 2: 257.

Mouille, G., Maddelein, M.-L., James, MG., and Ball, SG. (2000). Recent progress toward understanding the biosynthesis of the amylopectin crystal. Plant Physiology. 122: 989-997.

Nakamura, Y., Umenmotoo, T., Ogata, N., Kuboki, Y., Yano, M., and Sasaki, T. (1996). Starch debranching enzyme (R-enzyme or pullulanase) from developing rice endosperm: purification, cDNA and chromosomal localization of the gene. Planta. 199: 209-218.

Nelson, KE., Paulsen, IT., Heidelberg, JF., and Fraser, CM. (2000). Status of genome projects for nonpathogenic bacteria and archaea. Nat. Biotechnology. 18: 1049-1054.

Vrinten, PL., Nakamura, T. (2000). Wheat granule-bounc starch synthase I and II are encoded by separate genes that are expressed in different tissues. Plant Physiology. 122: 255-263.

Philippe, H., Douady, CJ. (2003). Horizontal gene transfer and phylogenetics. Current Opinion in Microbiology. 6: 498-505.

Preiss, J. (1982). Regulation of the biosynthesis and degradation of starch. Annual Review of Plant Physiol. 33: 431-454.

Raven, JA., Allen, JF. (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants?. Genome Biology. 4: 209

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., et al. (2000). Genome sequences of *Chlamydia trachomatis* MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Research. 28: 1397-1406.

Royo, J., Gimez, E., and Hueros, G. (2000). CMP-KDO synthetase: A plant gene borrowed from Gram-negative eubacteria. Trends in Genetics. 16: 432-433.

Schleucher, J., Vanderveer, P.J., Sharkey, T.D. (1998). Export of carbon from chloroplasts at night. Plant Physiology. 118: 1439-1445.

Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S., et al. (2000). Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. Nucleic Acids Research. 28: 2311-2314.

Shrager, J., Hauser, C., Chang, C-W., Harris, E.H., Daview, J., McDermott, J., Tamse, R., Zhang, Z., Grossman, A.R. (2003). *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of cDNA information. Plant Physiology. 131: 401-408.

Smith, AM., Denyer, K., Martin, C. (1997). The synthesis of the starch granule. Annual Review of Plant Physiology. Plant Molecular Biology. 48: 67-87.

Smith, A.M. (2001). The biosynthesis of starch granules. Biomacromolecules. 2, 335-341.

Stephens, RS., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, RL., Zhao, Q., et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science. 282: 754-759.

Thompson, JD., Higgins, DG., Givgon, TJ. (1994). ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position, specific gap penalties, and weight matrix choice. Nucleic Acids Research. 22: 4673-4680.

Tippman, H-F. (2004). Software review: Analysis for free: Comparing programs for sequence analysis. 5:82-87.

Trethwey, RN., Smith, AM. (2000). Starch metabolism in leaves. In: Leegood RC, Sharkey TD, von Caemmerer S, eds. Advances in photosynthesis. Vol. 9. Photosynthesis: physiology and metabolism. Dordrecht: Kluwer Academic Publishers. 205-232.

Turmel, M. et al. (1999). The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: insights into the architecture of ancestral chloroplasts genomes. Proceedings of the National Academy of Sciences. 96: 10248-10253

Wolf, YI., Aravind, L., Koonin, and EV. (1999). *Rickettsiae* and *Chlamydiae*: Evidence of horizontal gene transfer and gene exchange. Trends in Genetics. 15:173-175.

Yoo, S., Spalding, MH., Jane, J. (2002). Characterization of cyanobacterial glycogen isolated from the wild type and from a mutant lacking of branching enzyme. Carbohydrate Research. 337: 2195-2203.

Yoo, S., Moon, Y., Spalding, H., Jane, J. (in preparation). Insertional mutagenesis of glycogen synthase genes in cyanobacterium *Synechocystis* sp. PCC6803.

Yu, Y., Mu, HH., Mu-Forster, C., and Wasserman, BP. (1998). Polypeptides of the maize amyloplast stroma: Stromal localization of starch-biosynthetic enzymes and identification of an 81-kilodalton amyloplast stromal heat-shock cognate. Plant Physiology. 116: 1451-1460.

Zeeman, SC., Umenmoto T., Lue, W., Yeung, P., Martin, C., Smith, AM., Chen J. (1998). A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch. The Plant Cell. 10: 1699-1711.